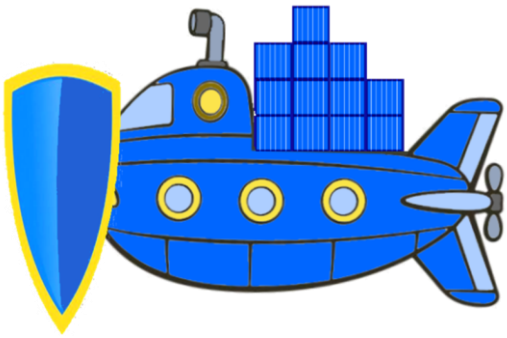


CyberNEMO

End-to-end CYBERsecurity to NEMO meta-OS

D5.1 Pilots Setup and Data Management Plan (DMP) v1

STATUS	FINAL	DUE DATE	31/03/2025
VERSION	1.0	DELIVERY DATE	31/03/2025

WORKPACKAGE	WP5	DOCUMENT	D5.1
LEAD PARTNER	ASM	DISSEMINATION LEVEL	PU
CONTRIBUTING PARTNERS	ASM, XGL, ENG, INTRA, SPACE, OTE, ESOFT, MUP, SU	REVIEWING PARTNERS	SYN CSL
KEYWORDS	Pilot setup, DMP, Data, Trial site, Living Labs, Open source		

© Copyright by the CyberNEMO Consortium

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101168182.

DISCLAIMER

This document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of its content.
This document may contain material, which is the copyright of certain CYBERNEMO consortium parties, and may not be reproduced or copied without permission. All CYBERNEMO consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.
Neither the CYBERNEMO consortium as a whole, nor a certain party of the CYBERNEMO consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered using this information.

	Participant organisation name	Short	Country
01	SYNELIXIS SOLUTIONS SA	SYN	EL
02	DNV AS	DNV	NO
03	THALES SIX GTS FRANCE SAS	TSG	FR
04	ENGINEERING - INGEGNERIA INFORMATICA SPA	ENG	IT
05	NETCOMPANY-INTRASOFT SA	INTRA	LU
06	SIEMENS SRL	SIEM	RO
07	MAGGIOLI SPA	MAG	IT
08	SPACE HELLAS SA	SPACE	EL
11	UNPARALLEL INNOVATION LDA	UNP	PT
12	TELEFONICA INNOVACION DIGITAL SL	TID	ES
13	OTE SA	OTE	EL
14	ASM TERNI SPA	ASM	IT
15	ENTERSOFT SA	ESOFT	EL
16	MEDITCINSKY UNIVERSITET-PLOVDIV	MUP	BG
17	SORBONNE UNIVERSITE	SU	FR
18	UNIVERSIDAD POLITECNICA DE MADRID	UPM	ES
19	RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN	RWTH	DE
20	PANEPISTIMIO PATRON	UPAT	EL
21	UNIVERSIDAD CARLOS III DE MADRID	UC3M	ES
22	ECLIPSE FOUNDATION EUROPE GMBH	ECL	DE
23	SPHYNX TECHNOLOGY SOLUTIONS AG	STS	CH
24	CYBERSOCIAL LAB SRL IMPRESA SOCIALE	CSL	IT
25	XGILITY LIMITED	XGL	IE

ACKNOWLEDGEMENT

This document is a deliverable of CYBERNEMO project. This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101168182.
The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

Document History

Version	Date	Contributor(s)	Description
V0.1	07/01/2025	Mohammad Ghoreishi, Francesca Santori (ASM)	Table of Contents (ToC)
V0.2	16/01/2025	Mohammad Ghoreishi (ASM)	First iteration after ToC
V0.3	20/01/2025	Mohammad Ghoreishi (ASM)	Trial set-up
V0.4	11/02/2025	Mostafa Jabari (ASM)	First round Trial contribution
V0.5	28/02/2025	Maria Belesioti (OTE), Nadia Khan (XGL), Antonis Gonos (ESOFT)	Second round Trial contribution
V0.6	04/03/2025	Javier Serrano (UPM)	Trial contribution
V0.7	14/03/2025	Mohammad Ghoreishi, Francesca Santori (ASM)	Ready for internal review
V0.8	26/03/2025	Mohammad Ghoreishi (ASM), Artemis Voulkidis (SYN), Luigi Briguglio (CSL)	Final version with internal reviewers and ready for quality check
V0.9	28/03/2025	Andreas Papadakis, Artemis Voulkidis (SYN)	Quality check
V0.9	31/03/2025	Dimitrios Skias (INTRA)	SAB check
V1.0	31/03/2025	Mohammad Ghoreishi, Francesca Santori (ASM), Andreas Papadakis, Artemis Voulkidis (SYN)	Final document

Document Reviewers

Date	Reviewer's name	Affiliation
14/03/2025	Artemis Voulkidis	SYN
20/03/2025	Luigi Briguglio	CSL

Table of Contents

DISCLAIMER 2
ACKNOWLEDGEMENT 2
Document History 3
Document Reviewers 3
Table of Contents 4
List of Figures 7
List of Tables 7
List of Acronyms and Abbreviations 8
Executive Summary 9
1. Introduction 10
1.1. Intended Audience 10
1.2. Relations to other activities 10
1.3. Document overview 11
2. Data Management Strategy 12
2.1. Introduction 12
2.1.1. Purpose of the Data Management Plan 12
2.1.2. Alignment with FAIR Principles and Open Science Practices 12
2.1.3. CyberNEMO-Specific DMP Context and Scope 13
2.1.4. Regulatory and Ethical Compliance 13
2.2. Data Summary 14
2.2.1. Overview of Data Types and Formats 14
2.2.2. Purpose of Data Collection and Reuse 14
2.2.3. Provenance of Data 15
2.2.4. Estimated Data Volume and Structure 15
2.2.5. Data Utility and Stakeholders 15
2.3. FAIR Data Management 15
2.3.1. Making Data Findable 15
2.3.2. Making Data Accessible 16
2.3.3. Making Data Interoperable 16
2.3.4. Ensuring Data Reusability 17
2.4. Data Security and Ethical Considerations 17
2.4.1. Cybersecurity Measures for Data 17
2.4.2. Anonymization, Pseudonymization, and Privacy Safeguards 17
2.4.3. Compliance with GDPR and Ethical AI Guidelines 18
2.4.4. Risk Management in Data Collection and Sharing 18
2.5. Open Science and Knowledge Sharing 19

2.5.1. Open Access to Publications 19
2.5.2. Sharing of Open Data and Models 19
2.5.3. Promoting Collaboration and Co-Creation 19
2.6. Research Outputs Beyond Data 20
2.6.1. Promoting Collaboration and Co-Creation 20
2.6.2. Certification Roadmap for CyberNEMO Compliance 20
2.6.3. Contribution to EU Meta-OS Projects 20
2.7. Allocation of Resources 21
2.7.1. Costs Associated with Data Management 21
2.7.2. Resource Allocation for Storage and Curation 21
2.7.3. Roles and Responsibilities in DMP Execution 21
2.8. Data Storage and Preservation 22
2.8.1. Long-Term Storage Infrastructure and Repositories 22
2.8.2. Retention Periods and Archiving Policies 22
2.8.3. Sustainability of Storage Solutions and Costs 22
2.9. Monitoring and Updates to the DMP 22
2.9.1. Iterative Updates Based on CyberNEMO Validation Processes 22
2.9.2. Integration of Stakeholder Feedback 23
3. Living Lab trials set-up and description 24
3.1. Trial #1: CyberNEMO Integration Infrastructure Technology Lab 25
3.1.1. Trial description and objective 25
3.1.2. Infrastructure and setup process 25
3.1.3. Dataset reuse strategy 25
3.2. Trial#2: Smart Energy & Smart Water Critical Infrastructures 25
3.2.1. Trial description and objective 25
3.2.2. Required Equipment/Assets 26
3.2.3. Data collection 27
3.3. Trial#3: Secure and Intelligent Media Content Supply Chain Ecosystems 32
3.3.1. Trial description and objective 32
3.3.2. Required Equipment/Assets 33
3.3.3. Data collection 34
3.4. Trial#4: Proactive & Intelligent Protection of Healthcare Critical Infrastructures 43
3.4.1. Trial description and objective 43
3.4.2. Required Equipment/Assets 43
3.4.3. Data collection 45
3.5. Trial#5: Smart Farming and FinTech/Logistics Supply Chain 45
3.5.1. Trial description and objective 45
3.5.2. Required Equipment/Assets 46

3.5.3. Data collection 47
3.6. Trial#6: Pan-European SAAM multi-Living Labs Federation 50
3.6.1. Trial description 50
4. Conclusion and Future Steps 51
5. References 52

List of Figures

Figure 1. Data types in CyberNEMO 14
Figure 2. Data source in CyberNEMO 15
Figure 3. Proprietary format adoption in CyberNEMO 17

List of Tables

Table 1. CyberNEMO Technologies and trials Federation 24
Table 2. Trials execution timeline 25
Table 3. Detailed description of equipment/assets in Trial 2. 26
Table 4. Trial 2 data collection 27
Table 5. DMP for Trial 2 datasets 28
Table 6. Detailed description of equipment/assets in Trial 3. 33
Table 7. Trial 3 data collection 34
Table 8. DMP for Trial 3 datasets 34
Table 9. Detailed description of equipment/assets in Trial 4 44
Table 10. Detailed description of equipment/assets in Trial 5. 46
Table 11. Trial 5 data collection 47
Table 12. DMP for Trial 5 datasets 48

List of Acronyms and Abbreviations

AI	Artificial Intelligence
AI HLEG	Artificial Intelligence High-Level Expert Group
AMI	Advanced Metering Infrastructure
CC	Computing Continuum
CDN	Content Delivery Network
EC	European Commission
EN	Enrichment Engine
EOSC	European Open Science Cloud
EVCS	Electric Vehicle Charging Station
DER	Decentralized Energy Resource
DMO	Data Manager Officer
DMP	Data Management Plan
DOI	Digital Object Identifier
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
IOT	Internet of Things
JSON	JavaScript Object Notation
LL	Living Lab
NHIF	National Health Insurance Fund
NHIS	National Health Information System
PMU	Phasor Measurement Unit
PQA	Power Quality Analyser
PV	Photovoltaic
SM	Smart Meter
SPCP	Situation Perception and Protection
UC	Use Case
VCE	Virtual Compression Engine
WP	Work Package
XML	Extensible Markup Language
ZTNA	Zero Trust Network Access

Executive Summary

The purpose of deliverable D5.1 is to present both the trial setup and the Data Management Plan (DMP) for the CyberNEMO project, supporting its mission to enhance cybersecurity and trust across the IoT-Edge-Cloud-Data Computing Continuum.
The Data Management Plan defines the methodology to collect, organize, anonymize, and share data in accordance with Horizon Europe's Open Research Data guidelines and the General Data Protection Regulation (GDPR). It covers IoT devices, edge-cloud infrastructures, CPU and network metrics, and IT equipment (e.g., gateways and firewalls), ensuring a robust and ethical data governance framework across the project lifecycle.
The document outlines the pilot setup across different Living Labs, describing the required hardware, environments, and trial-specific configurations. These setups serve as the foundational environments for testing and validating the CyberNEMO technologies under real-world conditions. Each pilot focuses on a distinct application domain (e.g., Smart Energy, Healthcare, Secure Media), and includes specific equipment such as sensors, controllers, network interfaces, and cyber-defence mechanisms. The configuration of these pilot environments directly informs the data lifecycle — including the types of data collected, the frequency of collection, and the integration of security and privacy measures from the outset.
The main goals of the deliverable are as follows:

- To detail the methodology for handling data and project outcomes in compliance with Horizon EU guidelines and GDPR.
- To identify data types to be shared as Open Data, including IoT metrics, edge-cloud characteristics, and anonymized operational insights.
- To outline the requirements for pilot adaptations, additional equipment procurement, and testbed setup, ensuring effective testing and validation processes.
- To define objectives and expected outcomes for each pilot, including Key Performance Indicators (KPIs) and data collection goals aligned with the CyberNEMO validation strategy.

CyberNEMO provides open access to:

- Scientific publications generated by the project.
- Open-source code for the CyberNEMO platform and associated tools.
- Anonymized datasets from pilot testing and validation activities.
- Technical documentation and white papers detailing project outcomes.

The project's data management approach balances transparency and openness with strict adherence to cybersecurity and privacy regulations. The DMP integrates innovations such as Zero-Trust Network Access, AI-explainable Situation Perception and Protection (SPCP) frameworks, and the SAAM collaborative platform for risk assessment, threat analysis, and incident mitigation.
By following the strategy laid out in this document, CyberNEMO aims to ensure secure, reproducible, and high-impact results through carefully orchestrated pilot activities and a structured, compliant approach to data governance.

1. Introduction

The CyberNEMO project advances cybersecurity and trust across the IoT-Edge-Cloud-Data Computing Continuum through a combined approach of rigorous data governance and practical, real-world validation. This deliverable presents a) the Data Management Strategy and the trial setup aspects of the project, recognizing their joint contribution to achieving the project's goals.
On one hand, the Data Management Strategy follows the Open Research Data Pilot in Horizon Europe to ensure effective, ethical, and secure handling of data throughout and beyond the project lifecycle. On the other side, the Living Lab trials provide the environments in which CyberNEMO technologies are deployed, tested, and validated under realistic operational conditions. These pilots directly inform the types of data collected, the security requirements, and the customization of data handling strategies.
This dual-track approach is central to CyberNEMO's mission: field validation ensures relevance and applicability, while a robust DMP guarantees compliance, openness, and data integrity.
This process involves:

- FAIR Principles: A methodology to ensure that research data generated within the CyberNEMO project is Findable, Accessible, Interoperable, and Reusable.
- Data Management Lifecycle: The identification, classification (i.e., open or confidential), organization, curation, preservation, storage, and sharing of data collected, processed, and/or generated during the project — including trial-specific data streams.
- Ethics and Legal Compliance: CyberNEMO applies GDPR-compliant data processing practices, as outlined in CyberNEMO deliverables D1.1 and D7.1, including risk assessments, consent management, and anonymization. These ensure ethical and legal alignment with EU and national regulations [1, 2].

1.1. Intended Audience

The deliverable is intended for a diverse range of stakeholders involved in both the pilot setup and the Data Management Plan (DMP) of the CyberNEMO project. On the pilot setup side, the primary audience includes technical partners and pilot site operators responsible for deploying and maintaining the testing environments, as well as technology developers who leverage pilot feedback to refine solutions, validate cybersecurity mechanisms, and tailor components for real-world scenarios. Additionally, domain-specific stakeholders—such as utilities, municipalities, media providers, and healthcare institutions—are critical actors engaged in each trial, while requirements engineers and system architects rely on insights from these setups to guide technical development and ensure contextual relevance.
On the DMP side, the audience encompasses consortium members such as researchers and data managers who implement data governance practices within the project, along with academic and research communities interested in cybersecurity, IoT, Edge-Cloud systems, and FAIR data principles. Industry stakeholders, including integrators and operators of critical infrastructure, are addressed, as are policymakers, regulatory bodies, and legal experts concerned with GDPR compliance and responsible data handling. The DMP is also highly relevant to open data advocates and the broader open-source community, ensuring transparency, accessibility, and reuse of the project's outputs. Furthermore, the CyberNEMO advisory board and external evaluators will use this deliverable to assess the alignment of pilot and data strategies with Horizon Europe guidelines and innovation objectives.
This deliverable represents the first version of the Pilot Setup and Data Management Plan. Future iterations, notably D5.2 scheduled for Month 18, will build upon this foundation by incorporating feedback from pilot execution, updated equipment configurations, and evolving requirements in data governance and cybersecurity validation as the CyberNEMO project progresses. This broad outreach ensures that both the pilot activities and data management practices are effectively implemented, validated, and exploited across multiple domains.

1.2. Relations to other activities

In relation to WP5, the DMP supports enabling systematic validation and penetration testing across different pilot environments. The DMP and pilot setup activities under Task 5.1 are integral to the broader goals of the

CyberNEMO project and directly support other project tasks. The DMP ensures efficient data collection, processing, anonymization, and sharing across all work packages, focusing on data from IoT devices, edge-cloud systems, and IT equipment (e.g., gateways and firewalls). It also guides the adaptation and customization of pilot environments to refine and validate the CyberNEMO platform under real-life conditions. Future updates of this deliverable (e.g., D5.2 in M18) will expand on this foundational version by integrating feedback from pilot deployments, evolving technology configurations, and lessons learned during the validation phase. Moreover, the DMP supports Task 5.2 by facilitating real-time data management for cybersecurity assessments in Smart Energy and Smart Water infrastructures. In Task 5.3, the DMP ensures structured data handling for secure and privacy-preserving multimedia content distribution. Similarly, Task 5.4 relies on the DMP for managing cybersecurity and privacy threats within Smart Healthcare Critical Infrastructures. In Task 5.5, the DMP ensures traceability and secure data management within the Smart Farming supply chain validation. Finally, Task 5.6 benefits from the DMP in enabling cross-pilot federation validation, ensuring GDPR compliance, secure data governance, and interoperability across multiple pilot sites.
In relation to WP1 (End-to-end CC Risks Assessment & Incidents Mitigation Strategies), the DMP supports risk analysis and threat modelling (Task 1.2) by ensuring structured data collection and continuous monitoring. It also aligns with Task 1.3 by ensuring compliance with GDPR, ethical, and legal requirements for data handling and sharing. Moreover, together with Task 1.3, it has been defined and submitted a baseline assessment questionnaire (for further details, see D1.1) to gather and collected information on data types and management. Specific DPM insights are reported in this document. Furthermore, the insights from DMP contribute to Task 1.4, which defines the federated Zero Trust architecture by providing data-driven validation and ensuring secure integration of computing continuum components. Last but not least, this document follows the ethics and regulatory guidelines defined in the "Project Management Handbook (D7.1) as a preliminary result of WP7 activities.

1.3. Document overview

The deliverable is structured as follows:

- Section 1 provides the introduction of the deliverable.
- Section 2 outlines the CyberNEMO data management strategy, providing a detailed framework for data collection, handling, security, and sharing, while ensuring compliance with FAIR principles, ethical guidelines, and open science practices.
- Section 3 provides an overview of the trial setup (mainly in terms of equipment) and the datasets identified during the early phases of the CyberNEMO project, organized by trial descriptions, objectives, required equipment, data collection processes, and execution timelines.
- Section 4 presents recommendations and further step.

2. Data Management Strategy

2.1. Introduction

2.1.1. Purpose of the Data Management Plan

A Data Management Plan (DMP) is a crucial aspect of effective data management, outlining the lifecycle of data collected, processed, and generated within a Horizon Europe project. As part of making research data findable, accessible, interoperable and re-usable (FAIR)1, a DMP should include information related to:

- The handling of research data during and after the end of the related project.
- What data will be collected, processed and/or generated.
- Which methodology and standards will be applied.
- Whether data will be shared/made open access.
- How data will be curated and preserved (including after the end of the related project).

The main requirements of the CyberNEMO data management strategy are listed as follows:

- Develop (and update) a Data Management Plan.
- Deposit data in a research data repository.
- Make possible for third parties to access, mine, exploit, reproduce and disseminate data – free of charge.
- Provide information on the tools and instruments needed to validate the results (or provide the tools).

The DMP needs to be updated during the project lifetime whenever significant changes arise, such as (non-exhaustive list):

- New data.
- Changes in consortium policies (e.g. new innovation, patent).
- Changes in consortium composition (e.g. addition of new consortium members).

The Data Management Plan for the CyberNEMO project outlines a comprehensive framework for managing data throughout the project's lifecycle. It serves to ensure that data generated, collected, processed, or reused under the CyberNEMO project activities adheres to the highest standards of accessibility, interoperability, and reusability [3]. Specifically, the DMP aims to:

- Facilitate the organization, documentation, and preservation of data to support reproducibility and transparency in research.
- Promote the sharing and reuse of datasets in compliance with Horizon Europe's Open Science principles.
- Address the unique challenges of managing data within the context of cybersecurity, IoT systems, and edge-cloud architectures.
- Ensure alignment with ethical, legal, and regulatory frameworks, particularly in relation to data privacy and protection.

2.1.2. Alignment with FAIR Principles and Open Science Practices

The CyberNEMO DMP aligns with the FAIR principles [4, 5], which emphasize making data Findable, Accessible, Interoperable, and Reusable. These principles are operationalized in line with the Horizon Europe guidelines

[1] Wilkinson, M.D., Dumontier, M., et al. (2016, March) The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. Available at: http://www.nature.com/articles/sdata201618.

outlined in [6], which recommend assigning persistent identifiers, applying standardized metadata, storing data in trusted repositories, and using community-endorsed standards and ontologies to ensure long-term usability and interoperability. Key actions to uphold these principles include:

- Assigning persistent identifiers and metadata to datasets to enhance discoverability.
- Storing data in trusted repositories to ensure long-term access and usability.
- Utilizing community-endorsed standards and ontologies to support data interoperability across diverse domains, including IoT, AI, and cybersecurity research.
- Providing clear licensing and documentation to facilitate the reuse of data by third parties.

In addition to FAIR principles, the DMP incorporates Open Science practices by [7]:

- Ensuring timely and open access to scientific publications and datasets.
- Sharing software, trained AI models, and experimental workflows via open repositories such as Zenodo and the European Open Science Cloud (EOSC).
- Supporting the reproducibility of research results through detailed documentation and open access to CyberNEMO's Living Labs and testing environments.

2.1.3. CyberNEMO-Specific DMP Context and Scope

The CyberNEMO project operates within the broader context of securing the IoT-edge-cloud continuum. The DMP reflects this unique focus by:

- Addressing the complexities of managing heterogeneous datasets, including IoT sensor data, edge computing metrics, AI/ML training datasets, and penetration testing results.
- Providing guidelines for anonymizing and processing sensitive data collected during the project's validation and testing phases.
- Enabling the integration of data management processes with CyberNEMO's innovative cybersecurity and AI/ML methodologies.
- Supporting cross-project collaboration by leveraging existing data and methodologies from related initiatives, such as IoT-NGIN [8] and NEMO [9].

The DMP's scope encompasses all data types generated or reused during the project, ensuring that both technical and ethical considerations are addressed in line with Horizon Europe requirements.

2.1.4. Regulatory and Ethical Compliance

CyberNEMO's data management practices are developed in alignment with applicable regulatory and ethical frameworks, with foundational guidance provided by the D7.1 "Project Management Handbook" and the D1.1 deliverable on ethics and legal requirements. These documents collectively define the procedures and responsibilities for ensuring lawful, transparent, and secure data handling throughout the project lifecycle. In addition, a dedicated Data Protection Officer (DPO) has been assigned to oversee GDPR compliance across all data management activities, supported by procedures for consent management, anonymization, and risk assessment. Specifically, the following frameworks inform the project's data governance:

- General Data Protection Regulation (GDPR): Ensuring that personal data is processed lawfully, transparently, and securely. Anonymization and pseudonymization techniques will be employed to safeguard privacy where applicable [1, 10]
- EU High-Level Expert Group on AI (AI HLEG) Guidelines: Adopting ethical AI principles to ensure that data handling aligns with human-centric and trustworthy AI practices.
- The EU AI Act [11]: Aligning with emerging legislation on the use of AI systems, particularly in areas related to cybersecurity, data privacy, and ethical considerations.

While not a central focus of the DMP, the Digital Operational Resilience Act (DORA)—addressing ICT risk management and operational continuity in critical sectors—has also been considered where relevant to the project's infrastructure and cybersecurity use cases. A more detailed analysis of such compliance is addressed in D1.1 (Ethics and Legal Requirements) and D7.1 (Project Management Handbook).

2.2. Data Summary

2.2.1. Overview of Data Types and Formats

The CyberNEMO project generates and processes a diverse range of datasets to support the development, validation, and assessment of cybersecurity solutions across multiple pilot environments. These datasets originate from real-world deployments in the project's Living Labs and cover IoT sensing, edge-cloud operations, AI/ML training, and system resilience evaluation. The types of data can be categorized as follows:

- IoT, Edge Cloud, and AI/ML Datasets:
  IoT datasets primarily consist of sensor readings from smart energy, water, health, media, and agricultural systems deployed in the CyberNEMO pilots. These include environmental data such as temperature, humidity, air quality, and energy consumption. Edge-cloud datasets capture system-level performance indicators such as latency, CPU usage, and bandwidth. AI/ML datasets include training inputs, intermediate outputs, model parameters, and validation results used to support cybersecurity mechanisms.
- Penetration Testing and Validation Data
  These datasets are collected during the evaluation of CyberNEMO's robustness against simulated cyberattacks. They include security logs, system responses, and performance under stress. Given their sensitive nature, the results of penetration testing are subject to strict access controls and are not publicly shared. Validation data also includes outcomes from system integration tests and user acceptance trials conducted within the pilots.

This specific context of the CyberNEMO project is reflected on the results of the baseline assessment carried out in collaboration with task T1.3, as shown in Figure 1. Datasets are mostly used in development, piloting and assessment. This type of data includes data from IoT, Edge Cloud, AI/ML, penetration testing and validation data.



Figure 1. Data types in CyberNEMO

2.2.2. Purpose of Data Collection and Reuse

Data collected in the CyberNEMO project serves multiple purposes, including validating cybersecurity measures, training and optimizing AI models, and demonstrating compliance with EU regulatory standards. The potential for data reuse beyond the project's lifecycle—particularly for non-sensitive datasets—will be assessed and detailed in Deliverable D5.2.

2.2.3. Provenance of Data

The datasets used in CyberNEMO originate from two primary sources: project-specific trials and data reused from related initiatives. The provenance of each dataset is carefully documented to ensure transparency and traceability.

Datasets generated during CyberNEMO trials are collected under controlled conditions to ensure alignment with the project's validation goals.

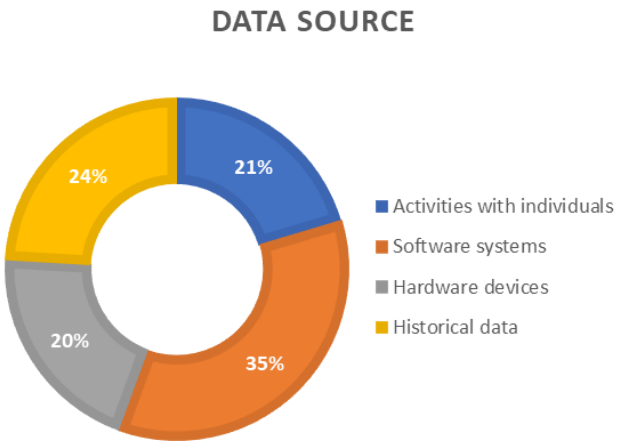


Figure 2. Data source in CyberNEMO

To build upon existing knowledge, CyberNEMO incorporates datasets from prior projects like NEMO and IoT-NGIN. These datasets are adapted to fit CyberNEMO's unique requirements while maintaining alignment with FAIR principles. The results of the baseline questionnaire confirm the provenance of data through the source of data, as shown in Figure 2.

Mainly, the data sources for piloting, testing and validation purposes represent the 71% of the CyberNEMO projects sources.

2.2.4. Estimated Data Volume and Structure

The volume of data generated during CyberNEMO activities is substantial, with an estimated size exceeding 200 GB. Structured data is organized into hierarchical formats for ease of analysis, while unstructured data such as video and image files are stored in scalable repositories optimized for large-scale retrieval.

2.2.5. Data Utility and Stakeholders

CyberNEMO data is valuable to a broad range of stakeholders, including researchers, policy makers, and industry practitioners. For researchers, the data provides a foundation for furthering knowledge in cybersecurity and IoT. For policy makers, it offers insights into the practical implementation of regulatory standards. Industry practitioners benefit from the data's applicability in developing and validating new technologies.

2.3. FAIR Data Management

2.3.1. Making Data Findable

CyberNEMO ensures that all datasets are findable by implementing defined metadata procedures and assigning persistent identifiers. Metadata will be structured using the Dublin Core standard, which has been adopted as the project-wide schema for describing datasets. A common metadata template will be developed under Task 5.1 to ensure consistency, and all data providers are required to populate it when preparing their datasets. This metadata will include essential fields such as dataset title, description, creator, creation date, format, license, and access rights. Persistent identifiers, such as DOIs, will be assigned upon publication through certified repositories (e.g., Zenodo) to ensure long-term traceability and reliable referencing.

- Metadata Standards and Persistent Identifiers: Metadata plays a critical role in documenting the origin, structure, and purpose of datasets. In CyberNEMO, metadata includes detailed descriptions of dataset content, collection methods, and contextual information. Persistent identifiers ensure long-term traceability and support interoperability between datasets.
- Naming Conventions and Discovery Optimization: Naming conventions are standardized across all datasets to enhance consistency and facilitate discovery. Datasets are named using a structured format that incorporates project identifiers, data type, and creation date. This approach supports efficient indexing and retrieval within repositories and search engines.

The data generated throughout the CyberNEMO project should be easily accessible to research communities. To ensure this, each dataset will be assigned a unique identifier.

CyberNEMO _[Name]_[Type]_[Date]_[Owner]

Where:

- [Name] is a short and characteristic name for the data
- [Type] is the type of data (code, publication, measured data)
- [Date] is the date when data was produced (format: DDMMYYYY)
- [Owner] is the owner (or owners) of the data (if exist)
- (underscore) is used as the separator between the fields.

2.3.2. Making Data Accessible

CyberNEMO data accessibility strategy ensures that datasets are available to authorized users while safeguarding sensitive information. Data is stored in trusted repositories like Zenodo and EOSC, which provide robust access controls and security features.

- Trusted Repositories and Access Protocols: All datasets are deposited in trusted repositories that adhere to international standards for data preservation and access. These repositories employ secure protocols to protect data integrity and enable controlled access for authorized users.
- Open Access vs. Restricted Access Policies: While most datasets are openly accessible, certain sensitive datasets are subject to restricted access policies. Access restrictions are based on legal, ethical, or security considerations and are clearly documented in the DMP.
- Embargoes and Access Committees: In cases where datasets are temporarily embargoed, clear timelines and conditions for lifting the embargo are established. Access committees are formed to evaluate requests for restricted datasets, ensuring compliance with project policies and regulatory standards.

2.3.3. Making Data Interoperable

Interoperability is a key focus of CyberNEMO data management strategy. Datasets and metadata are formatted according to widely accepted standards to facilitate integration with other systems and datasets.

- Standards for Data Formats and Vocabularies: CyberNEMO adopts standards such as JSON, XML, and CSV for data formats, and Dublin Core for metadata vocabularies. This approach ensures compatibility with existing tools and systems in the research community.
- Interdisciplinary Data Integration (IoT-Edge-Cloud Continuum): The project enables seamless data integration across IoT, edge, and cloud platforms by leveraging standardized APIs. This facilitates interoperability across disciplines and supports the development of complex, multi-domain applications.

For the sake of clarity, most of the data in CyberNEMO project (i.e., 91%) will use non-proprietary formats, as shown in Figure 3.

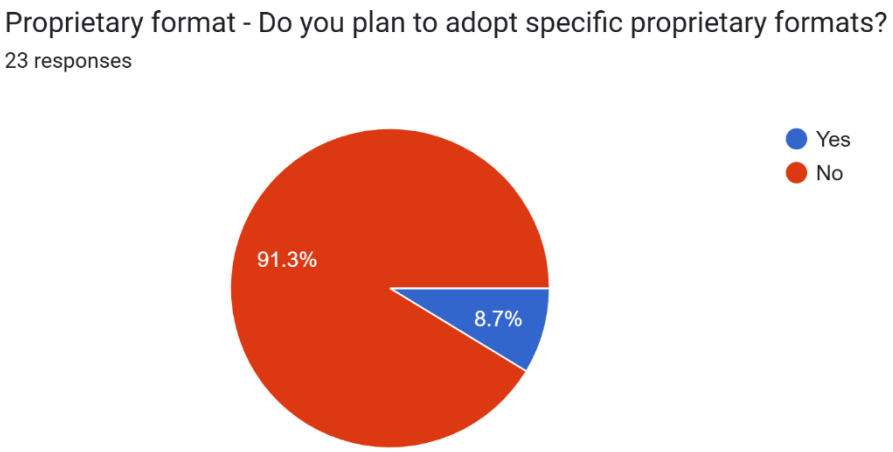


Figure 3. Proprietary format adoption in CyberNEMO

2.3.4. Ensuring Data Reusability

To maximize the reuse potential of datasets, CyberNEMO provides comprehensive licensing, metadata, and documentation. All reusable datasets are accompanied by user guides that describe their content, context, and intended applications to ensure clarity and reproducibility.

- Licenses and Documentation for Third-Party Use: Where applicable, datasets will be released under Creative Commons (CC) licenses to promote reuse while safeguarding intellectual property rights. However, sensitive datasets—such as penetration testing results or personally identifiable information—will be subject to access restrictions or internal use policies and will not be openly shared. Documentation will include methodology descriptions, variable definitions, data formats, and guidance on reuse.
- Provenance and Quality Assurance: CyberNEMO maintains detailed records of data provenance, including sources, collection tools, and processing workflows. Quality assurance is ensured through internal validation procedures, including automated consistency checks, peer review by task leaders, and metadata completeness verification prior to publication or sharing.

2.4. Data Security and Ethical Considerations

2.4.1. Cybersecurity Measures for Data

CyberNEMO adopts state-of-the-art cybersecurity measures to protect data integrity, confidentiality, and availability throughout its lifecycle. This includes the use of encrypted communication protocols (e.g., SSL/TLS) to secure data transmission and multi-factor authentication for access control. Intrusion detection systems are deployed within the pilot environments and core platform infrastructure to monitor network activity and detect unauthorized access attempts. Regular security audits and controlled penetration tests are conducted to identify and address vulnerabilities across the system components.

2.4.2. Anonymization, Pseudonymization, and Privacy Safeguards

To ensure compliance with data privacy regulations such as GDPR, all personal and sensitive data is anonymized or pseudonymized before storage or sharing. Anonymization techniques remove identifiable information to prevent re-identification, while pseudonymization replaces identifiable data with coded values. These measures are applied consistently across datasets to safeguard individual privacy while maintaining data utility for analysis. Further details are provided in D1.1, in terms of requirements for implementing technologies respecting fundamental rights (e.g., privacy) of individuals, aligned with the EU ethics and regulatory principles.

2.4.3. Compliance with GDPR and Ethical AI Guidelines

CyberNEMO strictly adheres to GDPR principles according to General Data Protection Regulation 2016/679 (Protection of personal data), ensuring that data is processed lawfully, transparently, and for legitimate purposes. Explicit consent is obtained from participants where applicable, and data minimization principles are observed to collect only necessary information. Ethical AI guidelines are integrated into the project's data management practices, ensuring that AI systems are designed and operated to respect human rights and avoid bias or discrimination.
The CyberNEMO consortium will abide by professional ethical practices and comply with ethics guidelines for trustworthy AI[2].
According to these Guidelines, trustworthy AI should be:

- lawful - respecting all applicable laws and regulations
- ethical - respecting ethical principles and values
- robust - both from a technical perspective, while considering its social environment

Further details are provided in D1.1, in terms of requirements for developing a Trustworthy AI, aligned with the EU ethics and regulatory principles.

2.4.4. Risk Management in Data Collection and Sharing

A robust risk management framework is implemented in CyberNEMO to address potential challenges related to data collection, processing, and sharing. To ensure secure and ethical data handling, the project follows a structured process for managing each dataset from its creation to its publication or restricted storage.

1. Characterization and Classification: Each dataset is first characterized by its origin, type (e.g., sensor data, AI models, penetration test logs), and sensitivity level (e.g., public, internal, confidential, or personal). This determines the applicable legal and ethical requirements.
2. Pre-processing and Anonymization: Datasets containing sensitive or personal information undergo pre-processing and anonymization using techniques aligned with GDPR. This step ensures that privacy risks are mitigated before any sharing or internal use.
3. Metadata and Documentation: Metadata is added based on a project-wide template to describe the dataset's content, structure, provenance, and applicable access conditions. This step supports traceability and accountability.
4. Risk Assessment and Approval: Before provision, datasets are reviewed as part of a regular risk assessment process involving cybersecurity leads and the Data Protection Officer (DPO). Approval is required for any public release.
5. Provision and Repository Selection: Depending on the dataset's classification, it is either uploaded to a public repository (e.g., Zenodo) under an appropriate license, or stored in a restricted-access repository for internal use only. Access control policies are enforced accordingly.

Additional safeguards are in place throughout the data lifecycle, such as training for involved personnel, periodic audits, and pre-defined response protocols for handling potential data breaches or ethical issues. This systematic approach helps ensure that data sharing remains compliant with legal, ethical, and cybersecurity standards, while also supporting transparency and stakeholder confidence.

[2] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

2.5. Open Science and Knowledge Sharing

2.5.1. Open Access to Publications

The European Commission has early recognized that research data is as important as publications[3]. In 2012 it was announced that it would experiment with open access to research data[4] aiming to facilitate broader and faster access to scientific papers and datasets. This would enable researchers and businesses to build upon publicly funded research more efficiently.
As an initial step, the Commission established open access to scientific publications. As of the year 2014, all articles produced with EU-funding had to be accessible according to the following options:

- Gold open access: Articles were immediately available online through the publisher with up-front publication costs eligible for reimbursement by the European Commission.
- Green open access: Researchers had to make their articles available through an open access repository within six months of publication (12 months for articles in the fields of social sciences and humanities).

Additionally, the Commission recommended that Member States adopt similar open-access policies for results of research funded under national programmes. This effort strengthens Europe's innovation capacity and ensures that citizens benefit from scientific advancements.

- Peer-Reviewed Articles and Open Science Repositories: CyberNEMO ensures that all peer-reviewed publications resulting from the project are accessible via trusted open science repositories, such as OpenAIRE, Zenodo, and the European Open Science Cloud (EOSC). Articles are made available under open access policies, enabling researchers, practitioners, and the public to benefit from the project's scientific contributions.
- FAIR Publication Metadata: To align with the FAIR principles, metadata for all publications is enriched with key descriptors such as author information, publication date, project identifiers, and funding details. This approach supports seamless discovery and integration into global scholarly ecosystems.

2.5.2. Sharing of Open Data and Models

- AI/ML Algorithms and Training Data: The project emphasizes the sharing of AI/ML models and training datasets through open-access repositories. These resources include algorithm documentation, parameter settings, and validation data to facilitate reuse and reproducibility in related research fields.
- Data Repositories (e.g., Zenodo, EOSC): All open datasets, including research outputs and supporting materials, are deposited in trusted repositories like Zenodo and EOSC. These platforms ensure long-term preservation, accessibility, and proper citation of project data, fostering transparency and collaboration.

2.5.3. Promoting Collaboration and Co-Creation

- Open Access to CyberNEMO Living Labs: CyberNEMO Living Labs offer a collaborative environment where project partners and selected external stakeholders can engage with the project's tools, use cases, and methodologies. Access is coordinated in alignment with pilot-specific conditions and project commitments, enabling opportunities for co-creation, validation, and feedback. This approach promotes innovation and supports the real-world application of project results.

[3] European Commission (2012, July). Communication on "Towards better access to scientific information: Boosting the benefits of public investments in research" [COM (2012) 401 final, 17.07.2012]. Available at: http://ec.europa.eu/research/cience-society/document_library/pdf_06/era-communication-towards-betteraccess-to-scientific-information_en.pdf

[4] European Commission (2012, July). Press Release IP/12/790 - Scientific data: open access to research results will boost Europe's innovation capacity. Brussels, July 2012. Available at: http://europa.eu/rapid/press-release_IP-12-790_en.htm

- **e-Trainings and Reproducibility Initiatives:** To enhance knowledge dissemination and capacity building, CyberNEMO organizes e-trainings and workshops focused on the project's technologies and methodologies. These initiatives are designed to improve the reproducibility of research and foster collaborative problem-solving among diverse stakeholders.

2.6. Research Outputs Beyond Data

2.6.1. Promoting Collaboration and Co-Creation

The project prioritizes the development and dissemination of open-source software and tools to ensure transparency and promote wide-scale adoption. These tools include:

- **AI/ML Modules:** Advanced algorithms for cybersecurity, trust validation, and privacy safeguards, shared under business-friendly licenses.
- **System Monitoring Tools:** Solutions for real-time performance tracking across IoT, edge, and cloud systems.
- **APIs and Integration Frameworks:** Modular APIs designed for interoperability within the IoT-edge-cloud continuum, facilitating seamless integration with existing systems.

All software is shared on platforms like GitHub and the Eclipse Foundation, ensuring accessibility and fostering community-driven improvements.

2.6.2. Certification Roadmap for CyberNEMO Compliance

To support the adoption and long-term sustainability of the CyberNEMO framework, the project explores the development of a certification roadmap aligned with relevant EU regulations and standards. This roadmap outlines possible paths towards the certification of CyberNEMO components and outcomes, based on risk classifications and applicable rules. These activities are supported primarily under Task 1.3, which defines the foundational compliance and trustworthiness framework for the project. The roadmap considers the following elements:

- **Standards for Cybersecurity and Privacy:** Detailing compliance with EU regulations such as GDPR and the AI Act.
- **Validation Protocols:** Criteria for ensuring robustness, trustworthiness, and ethical AI practices.
- **Industry Best Practices:** Recommendations for secure deployment and maintenance of IoT and AI solutions.

This roadmap enables organizations to perform certification paths of their systems according to the EU background regulatory framework.

2.6.3. Contribution to EU Meta-OS Projects

CyberNEMO actively contributes to the broader ecosystem of European meta-operating system projects, including collaborations with IoT-NGIN and NEMO. Key contributions include:

- **Enhanced Cybersecurity Protocols:** Sharing methodologies for zero-trust architectures and federated security models.
- **Data and Model Interoperability:** Providing tools and frameworks that support the seamless integration of data across projects.
- **Knowledge Exchange:** Participating in joint workshops and knowledge-sharing initiatives to harmonize efforts across related projects.

2.7. Allocation of Resources

2.7.1. Costs Associated with Data Management

The implementation of CyberNEMO Data Management Plan involves costs related to component supporting secure, ethical, and FAIR-compliant data handling. These include:

- Data Storage and Hosting: Expenses related to scalable infrastructure for short-term and long-term storage, including secure environments for sensitive data.
- Repository Fees: Potential costs for depositing datasets in trusted repositories, especially those offering advanced metadata support or long-term preservation.
- Data Anonymization and Security: Investments in tools and services to ensure compliance with GDPR and cybersecurity requirements, including anonymization techniques, encryption protocols, and access control systems.
- Personnel and Training: Funding for data management specialists, capacity-building activities, and staff training on FAIR principles, open science practices, and secure data handling.

2.7.2. Resource Allocation for Storage and Curation

To manage the large volumes of data generated in the project, resources are allocated strategically across several key areas:

- Storage Infrastructure: Resources are reserved for scalable and secure storage solutions, which include institutional data centres and trusted cloud environments capable of hosting diverse formats such as high-resolution IoT sensor data, system logs, and AI/ML models. Where feasible, existing infrastructure from partner organizations is utilized to optimize cost and performance.
- Data Curation Tools: Partners apply specialized tools and internal platforms for metadata creation, automated quality checks, and dataset cataloguing. These tools support adherence to FAIR principles and enable uniform documentation across datasets.
- Redundancy and Backup Systems: To ensure data resilience, redundant storage layers are included in institutional setups or cloud storage plans. Regular backups and integrity checks are standard operating procedures for sensitive or high-volume datasets.
- Long-Term Preservation: To ensure data availability beyond the project's lifecycle, CyberNEMO plans to deposit selected datasets in certified repositories that support persistent identifiers (e.g., Zenodo, institutional archives). Data requiring extended preservation is documented with access conditions and metadata to support long-term reuse and traceability

2.7.3. Roles and Responsibilities in DMP Execution

The successful execution of the DMP relies on a well-defined structure of roles and responsibilities:

- Data Manager Officers (DMOs): Assigned for each pilot site, responsible for overseeing data collection, quality assurance, and adherence to DMP protocols.
- IT and Security Teams: Ensure the implementation of cybersecurity measures, secure storage, and compliance with GDPR and other regulations.
- Project Partners: Contribute to the generation, curation, and sharing of datasets while maintaining compliance with ethical guidelines.
- DMP Monitoring Committee: Regularly reviews the DMP to address emerging challenges, update policies, and ensure alignment with Horizon Europe requirements.

2.8. Data Storage and Preservation

2.8.1. Long-Term Storage Infrastructure and Repositories

CyberNEMO ensures secure and scalable long-term storage for all data generated during the project. Key elements include:

- **Storage Infrastructure:** The project employs a combination of high-performance cloud platforms and institutional servers. These systems are designed to accommodate diverse data formats, including structured IoT datasets, unstructured video files, and AI/ML models.
- **Trusted Repositories:** Data is deposited in repositories such as Zenodo, EOSC, and 4TU.Research Data, all of which comply with international standards for preservation and accessibility.
- **Redundancy and Backup Systems:** Backup strategies are implemented to prevent data loss, including periodic mirroring of datasets and offsite storage for added resilience.

2.8.2. Retention Periods and Archiving Policies

The archiving policies of the project align with Horizon Europe's recommendations and ensure compliance with FAIR principles:

- **Retention Periods:** Most datasets are retained for a minimum of five years following the project's completion. Critical datasets, such as AI/ML training data and cybersecurity validation logs, may be retained longer for continued research and collaboration.
- **Archiving Policy:** After the active phase of the project, data transitions to cold storage systems to reduce costs while maintaining accessibility for authorized users. Personal data is securely anonymized or deleted in accordance with GDPR regulations.
- **Metadata Preservation:** Metadata associated with datasets is archived indefinitely to ensure long-term findability and accessibility.

2.8.3. Sustainability of Storage Solutions and Costs

To ensure the sustainability of storage solutions:

- **Scalable Systems:** Storage solutions are selected based on their ability to scale with growing data volumes and handle computational demands.
- **Cost Management:** Long-term costs are mitigated through the use of cost-effective repositories like EOSC and institutional support from project partners.
- **Institutional Support:** Participating institutions provide financial and technical support for data storage beyond the project's lifecycle, ensuring ongoing accessibility and preservation

2.9. Monitoring and Updates to the DMP

2.9.1. Iterative Updates Based on CyberNEMO Validation Processes

The DMP is treated as a dynamic document that evolves in response to project developments:

- **Validation-Driven Updates:** As CyberNEMO pilot trials progress, new insights inform updates to data management practices, including data classification, security protocols, and sharing mechanisms.
- **Annual Reviews:** A formal review process is conducted annually to assess the DMP's alignment with project objectives and FAIR principles.
- **Emerging Technologies:** The DMP incorporates advancements in data storage, cybersecurity, and AI/ML, ensuring that CyberNEMO remains at the forefront of data management innovation.

2.9.2. Integration of Stakeholder Feedback

Stakeholder engagement is integral to refining the DMP:

- Feedback Mechanisms: Regular consultations with consortium partners, external reviewers, and end-users provide actionable insights for improving data management strategies.
- Collaborative Updates: Stakeholder input is incorporated into DMP revisions to ensure the document reflects practical needs and addresses emerging challenges.
- Dissemination: Updated versions of the DMP are shared with all stakeholders to maintain transparency and ensure consistent implementation across the project.

To conclude, DMP represents a cornerstone of the project's commitment to Open Science and ethical data management. Its iterative approach ensures adaptability to new challenges and technologies while maintaining alignment with European standards. The project's robust data strategy not only supports its objectives but also sets a benchmark for future initiatives in the IoT-edge-cloud continuum. By prioritizing transparency, security, and collaboration, CyberNEMO ensures that its data outputs contribute meaningfully to societal and scientific advancements. Moreover, the DMP outlines a comprehensive strategy for managing the lifecycle of project data:

- Goals: Ensuring data is FAIR-compliant, securely stored, ethically managed, and widely accessible for reuse and collaboration.
- Roadmap: Detailed policies and practices are in place for data collection, storage, sharing, and preservation. These efforts align with Horizon Europe's Open Science principles and EU regulatory requirements.
- Impact: By leveraging advanced data management practices, CyberNEMO fosters innovation, enhances cybersecurity resilience, and contributes to the broader research ecosystem.

3. Living Lab trials set-up and description

This section introduces the trials setup, providing an overview of the trial sites. It includes a summary of where the Use Cases (UCs) and scenarios will be conducted, the actors involved, and the primary components or equipment required for each UC. This section presents a list of the necessary equipment along with their procurement status—indicating whether the equipment is already available or requires procurement. This encompasses various types of equipment, such as servers, sensors, vehicles, robots, drones, and software. Additionally, the section discusses the datasets that have been identified and, where applicable, already collected for the trials. These datasets play a pivotal role in supporting the execution and evaluation of the UCs.
A critical focus of this deliverable is the alignment of each UC with the CyberNEMO technologies. This section marks the initial collaboration between WP5 and the technical partners from WP2, WP3, and WP4 to identify the technologies involved in each trial use case.
To summarize, Table 1 provides an overview of the alignment between CyberNEMO technologies and the project's trial environments. Each technology component, categorized by its originating work package, is mapped to one or more use cases, ensuring that no component remains unutilized. Likewise, every trial integrates specific CyberNEMO technologies, confirming a federated and collaborative approach across pilots. This structured mapping supports cross-validation, shared infrastructure use, and the technological coherence of the overall project architecture.

Table 1. CyberNEMO Technologies and trials Federation

CyberNEMO technologies	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
Dynamically updated repository of cyber/procedural threats (R1)	✓	✓	✓	✓	✓	✓
Homomorphic encrypted ZT-FML for intelligence on device (R2)	✓	✓	✓	✓		✓
Cybersecure extensions to mNCC (R3) and ZTNA layer (R4)	✓		✓	✓		✓
Network Topology Attestation &Virtual Micro-segmentation (R5)	✓				✓	✓
Cybersecurity, trust and privacy (R6) and Extended SASE (R7)	✓	✓	✓	✓	✓	✓
Cyber threat intelligence (R8), Cybersecure SPCP (R9)		✓		✓	✓	✓
PRESS framework (R10) and SLO-based PPE (R11)		✓	✓	✓		✓
Evidence-based Certification & Accreditation (R12)	✓	✓				✓
Sharing, Assessment, Analysis and Mitigation (SAAM) (R13)			✓	✓	✓	✓
Culture of Cybersecurity and Living Labs (R14/R15)	✓	✓	✓	✓	✓	✓

The execution timeline of trials is divided into multiple phases and is detailed in Table 2 below.

Table 2. Trials execution timeline

Phase	Estimated start date	Estimated end date
Trials set-up and equipment procurement	M4	M8
Initial implementation and validation	M9	M18
Intermediate implementation and validation	M19	M30
Final implementation and validation	M31	M36

While each pilot addresses distinct use cases and domain-specific challenges, the IoT–Edge–Cloud Computing Continuum (CC) serves as a foundational layer across the CyberNEMO architecture. The degree of visibility and integration of the CC may vary depending on the specific trial setup, but it remains a unifying concept that underpins secure data flows and federated validation. Some aspects of the scenarios, including technical details, equipment deployment, and dataset availability, are expected to evolve in the next iteration of the deliverable as the trials mature.

3.1. Trial #1: CyberNEMO Integration Infrastructure Technology Lab

3.1.1. Trial description and objective

Trial #1 serves as the infrastructure foundation for the CyberNEMO technology lab, building upon the OneLab platform for cybersecurity, IoT, and 5G/6G integration. The datasets used in this trial will be derived from those indicated / created in other trials, ensuring comprehensive validation of CyberNEMO components and methodologies.
OneLab will be utilized to validate critical CyberNEMO technologies included in Table 1. This trial will leverage OneLab's state-of-the-art, virtualized, and programmable resources to create heterogeneous experimental AIoT cybersecurity scenarios, enabling the exploration of secure and resilient digital infrastructures.

3.1.2. Infrastructure and setup process

The setup process for Trial #1 involves deploying a shared testing environment within the OneLab facility, where components from other trials can be validated. This environment includes different types of Kubernetes clusters based on their purpose and technology. Standard Kubernetes clusters are used for development, production, and some staging environments, providing full orchestration features. Lighter K3s clusters are used in staging scenarios where simplicity and low resource usage are preferred. Each cluster is set up using automated scripts and configuration files, which helps ensure that the setup is consistent and repeatable across all environments. Network settings and resource allocations are customized to reflect real-world deployment conditions, making this environment a reliable platform for testing.

3.1.3. Dataset reuse strategy

Trial #1 serves as a testing environment where components from other trials are validated. Instead of introducing new datasets, it reuses the datasets defined in those trials. This ensures compatibility, streamlines testing processes, and maintains consistency across all components being evaluated.

3.2. Trial#2: Smart Energy & Smart Water Critical Infrastructures

3.2.1. Trial description and objective

ASM is responsible for Trial#2 to evaluate CyberNEMO effectiveness in protecting critical infrastructures through smart energy systems and smart water networks. The trial aims to evaluate CyberNEMO solutions to verify their ability to manage and stabilize the electricity grid while studying how cyberattacks impact the water network. ASM provides strong infrastructure support for Trial#2 through its Advanced Metering Infrastructure (AMI) and Smart Meters (SMs) logs, Photovoltaic (PV) plants, Electric Vehicle Charging Stations (EVCSs), water supply

infrastructure, Power quality analyser (PQAs), Phasor Measurement Units (PMUs), which are intended for the analysis of renewable energy sources. Remote-controlled operations allow for efficient distribution and control of consumer behaviour in electricity consumption and water supply network. The main goals consist of improving smart grid cybersecurity measures while safeguarding Decentralized Energy Resources (DERs) and reducing privacy risks from data breaches.
Use case applications:

- Smart Grid Cybersecurity: Smart Grid Cybersecurity protects power networks, substations, and SMs from cyber threats to ensure reliable electricity generation and consumption. It detects cyberattacks, system failures, and human errors in network control, including maintenance and firmware updates. It also analyses how cyberattacks on electricity infrastructure can impact water networks, which may act as energy storage. Key risks include SCADA system disruptions and coordinate cyberattacks that exploit dependencies between electricity and water systems. This ensures smart grids remain secure, stable, and resilient against evolving cyber threats.
- Privacy and data breaches: Privacy and data breaches in smart grids involve protecting energy consumption data from unauthorized access. Analysing anonymized SM and systems data helps detect unusual energy demand while keeping individual usage and behaviour private. Strong privacy measures prevent cyber threats targeting energy producers, consumers, EVs, and distributed energy resources (DERs). Risks include hacking smart meters, cyberattacks on energy networks, and threats to EV charging systems. Ensuring data privacy keeps smart grid operations secure and protects user information from breaches.

3.2.2. Required Equipment/Assets

Table 3 below shows a detailed description of the required equipment for the scenarios, along with its type and status (whether it is available or still to be procured).

Table 3. Detailed description of equipment/assets in Trial 2.

Scenario	Scenario description	Equipment/Asset	Type	Status
Scenario 1: Data Breach via SMs & pfSense firewalls	This scenario focuses on protecting SM data and AMI from cyber threats and unauthorized access. It analyzes pfSense firewall logs and network security data to detect possible data breaches, privacy risks, and suspicious network activity.	AMI	IoT	Available
		SMs	IoT	Available
		Consumption data	Cloud platform(database)	Available
		pfSense firewall	IoT	Available
Scenario 2: Data Breach & Ransomware on EV Charging Stations & PV Plants	This scenario focuses on securing EVCS and PV plants from cyber threats, including	PV	IoT	Available
		EVCS	IoT	Available
		Consumption data	Cloud platform(database)	Available

	data breaches, ransomware attacks, and unauthorized access.	Production data	Cloud platform(database)	Available
		WatchGuard tool	IoT	Available
		Darktrace tool	IoT	Available
Scenario 3: PQA Data Manipulation for Distribution Grid & Water Supply	This scenario involves cyberattacks targeting PQA data to manipulate power quality measurements and water data readings.	PQAs	IoT	Available
		Water supply infrastructure	IoT/Cloud platform(database)	Available
		Electricity substations and transformer	IoT/Edge	Available
Scenario 4: Privacy & Cyber Threats via PMUs in a Distribution Feeder	This scenario focuses on securing PMU data from cyber threats, including false data injection, unauthorized access. Attackers may manipulate PMU data to disrupt grid stability.	PMUs	Edge	M9
		Electricity substations and transformer	IoT/Edge	Available
		Network security systems	Edge	Available

3.2.3. Data collection

The following datasets have been identified as part of the use case:

Table 4. Trial 2 data collection

Dataset	Description/Specifications	Status
PMU	Measurement device in a portion of the grid for fault detection, state estimation, real-time visualization, etc.	M9
PQA	Electrical measurement device in secondary substations	M9
PV plants	ASM headquarters PV	M9
EVCS	Electric Vehicle Charging Station in ASM headquarters	M9
Smart meters	Real-time electrical measurement of residential end-users	M9
Water reservoir	Level of water, flow rate, and pressure in a water pump station	M9

AMI	Smart grid technology that enables two-way communication between utilities and consumers through smart meters	M9
Pfsense system log	Open-source firewall with system logs for monitoring	M9
WatchGuard	A commercial System Firewall in ASM	M9
Darktrace	Cybersecurity tool to detect, analyze, and respond to protect the network. It checks traffic on the entire ASM network and sends alerts to Sophos (anonymized)	M9

Table 5. DMP for Trial 2 datasets

Dataset #1	Dataset name: PMU
Brief Description	This is a device that measures phasors in the power grid, providing real-time voltage and current measurements for fault detection, state estimation, and grid monitoring.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The data will be transmitted using the MQTT protocol over a public IP and/or through LAN or VPN connections, ensuring fast, reliable, and secure communication for real-time data exchange.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Each PMU is assigned a unique ID, and its historical data is securely stored in an SQL database for efficient retrieval and analysis.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	The dataset will come with clear documentation, explaining its structure, data collection process, and how to use it for easy understanding and analysis.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data, along with data validation procedures and data cleaning techniques to ensure accuracy and reliability.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The data will be shared with the entire Project Consortium, and any use beyond the CyberNEMO Project must be approved by ASM.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The pilot technical leader will manage the dataset throughout the project's duration, with an external repository being considered for storage after the project ends.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The data is anonymized, and access to the database is restricted to specific IPs and secured with a password. Access to the MQTT Broker is also password-protected for enhanced security.

Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM is responsible for database maintenance, while the organization's Data Protection Officer ensures compliance with EU data protection laws.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data and ASM's management plan follow all national and European laws to ensure proper data protection and handling.
Dataset #2	Dataset name: PQA
Brief Description	This is an electrical measurement device used in secondary substations to monitor and analyse voltage, current, power, and quality parameters in real-time. It helps detect harmonics, voltage sags, frequency variations, and other power disturbances, ensuring a stable and efficient power supply.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data collection is performed each 1 minute through MQTT broker, using APIs, public IP, and direct database queries.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	PQA's have Unique ID and the historical data will be stored in an SQL database.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the dataset.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data, along with data validation procedures and data cleaning techniques to ensure accuracy and reliability.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The data will be shared with the entire Project Consortium, and any use beyond the CyberNEMO Project must be approved by ASM.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The pilot technical leader will manage the dataset throughout the project's duration, with an external repository being considered for storage after the project ends.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The data is anonymized, and access to the database is restricted to specific IPs and secured with a password. Access to the MQTT Broker is also password-protected for enhanced security.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM is responsible for database maintenance, while the organization's Data Protection Officer ensures compliance with EU data protection laws.

Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data and ASM's management plan follow all national and European laws to ensure proper data protection and handling.
Dataset #3	Dataset name: EVCS
Brief Description	The voltage, current, and power of the charging station are continuously monitored to ensure efficient operation, stable energy supply, and optimal charging performance for connected vehicles.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data is collected every 5 seconds using the OCPP protocol, which facilitates communication between charging stations and the central management system. The WebSocket protocol enables real-time, full-duplex communication over a single TCP connection. All collected data is non-personal and used solely for system monitoring and optimization.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Each charging station is assigned a unique ID, and its historical data is securely stored in an SQL database for tracking, analysis, and efficient management.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	All data is stored on the server and will be retained for five years after the project's completion to ensure long-term accessibility and compliance.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data, along with data validation procedures and data cleaning techniques to ensure accuracy and reliability.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The data will be shared with the entire Project Consortium, and any use beyond the CyberNEMO Project must be approved by ASM.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The pilot technical leader will manage the dataset throughout the project's duration, with an external repository being considered for storage after the project ends.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The data is anonymized, and access to the database is restricted to specific IPs and secured with a password. Access to the MQTT Broker is also password-protected for enhanced security.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM holds ownership of the charging station dataset, managing its storage, access, and usage in accordance with project guidelines and regulations.

Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data and ASM's management plan follow all national and European laws to ensure proper data protection and handling.
Dataset #4	Dataset name: SMs
Brief Description	Measurements of voltage, current, and power parameters are continuously monitored in ASM energy units to ensure efficient operation, stability, and performance analysis.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data will be transmitted using the MQTT protocol over a public IP or through LAN/VPN connections, ensuring secure and reliable communication.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Each Smart Meter has a unique ID, and its historical data is securely stored in an SQL database for tracking, analysis, and efficient management.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	The dataset will come with clear documentation, explaining its structure, data collection process, and how to use it for easy understanding and analysis.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data, along with data validation procedures and data cleaning techniques to ensure accuracy and reliability.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The data will be shared with the entire Project Consortium, and any use beyond the CyberNEMO Project must be approved by ASM.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The pilot technical leader will manage the dataset throughout the project's duration, with an external repository being considered for storage after the project ends.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The data is anonymized, and access to the database is restricted to specific IPs and secured with a password. Access to the MQTT Broker is also password-protected for enhanced security.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM is responsible for database maintenance, while the organization's Data Protection Officer ensures compliance with EU data protection laws.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data and ASM's management plan follow all national and European laws to ensure proper data protection and handling.
Dataset #5	Dataset name: Firewall and Network Security Log Management data

Brief Description	Firewall and Network Security Log Management involves collecting, storing, and analysing logs from security devices to monitor network activity, detect threats, and ensure compliance.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data will be transmitted as CSV file reports.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Each dataset has a unique ID, and its historical data is securely stored for tracking, analysis, and efficient management.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	The dataset will come with clear documentation, explaining its structure, data collection process, and how to use it for easy understanding and analysis.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data, along with data validation procedures and data cleaning techniques to ensure accuracy and reliability.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The data will be shared with the entire Project Consortium, and any use beyond the CyberNEMO Project must be approved by ASM.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The pilot technical leader will manage the dataset throughout the project's duration, with an external repository being considered for storage after the project ends.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The data is anonymized, and access to the database is restricted to specific IPs and secured with a password. Access to the MQTT Broker is also password-protected for enhanced security.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM is responsible for database maintenance, while the organization's Data Protection Officer ensures compliance with EU data protection laws.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data and ASM's management plan follow all national and European laws to ensure proper data protection and handling.

3.3. Trial#3: Secure and Intelligent Media Content Supply Chain Ecosystems

3.3.1. Trial description and objective

This trial under CyberNEMO aims to validate end-to-end cybersecurity measures for professional media workflows spanning contribution, production, and distribution. By integrating zero-trust approaches with advanced technologies (e.g., ZTNA, ZT-FML, IPDM-DSS), the trial demonstrates proactive threat mitigation, data protection, and the reliable handling of multimedia content within a multi-domain, multi-cluster Edge-Cloud

environment. Objectives include identifying vulnerabilities across IoT, Edge, and Cloud components, ensuring integration with European cybersecurity standards, and quantifying performance in threat detection, incident response, and data protection.
Use case applications:

- Collaborative and Secure Multimedia Content Production: This scenario focuses on a media factory environment where multiple creators and production teams, often geographically dispersed, collaborate on multimedia assets. The workflow emphasizes security at each stage of content contribution and production, leveraging technologies like end-to-end encryption, zero-trust network access, and digital signatures. By ensuring identity verification and robust monitoring, the system guards against potential threats such as data interception, unauthorized access to creative assets, and content manipulation, thus preserving the integrity and confidentiality of all media resources.
- Efficient and Secure Multimedia Content Distribution Across Multi-Domain Edge-Cloud: In the second scenario, the trial addresses the secure delivery of high-quality media content through distributed content delivery network (CDN) nodes. This scenario highlights the need for high availability and low latency, achieved through dynamic resource allocation and advanced load-balancing strategies. Simultaneously, the implementation of tools like AI-based firewalls, real-time traffic monitoring, and token-based authentication helps detect and mitigate threats such as DDoS attacks, cache poisoning, and unauthorized content access. As a result, the system delivers optimized performance and robust security, ensuring end users receive timely, trustworthy media even under variable network conditions.

3.3.2. Required Equipment/Assets

Table 6 below shows a detailed description of the required equipment for the scenarios, along with its type and status (whether it is available or still to be procured).

Table 6. Detailed description of equipment/assets in Trial 3.

Scenario	Scenario description	Equipment/Asset	Type	Status
Multimedia Content Production	Ensures secure collaboration across distributed media production environments, safeguarding content authenticity and defending against cyberthreats.	Feeder	Edge	Available
		Media Production Engine (MPE)	Edge	Available
		Virtual Compression Engine (vCE)	Edge/Cloud	Available
		Enrichment Engine (EN2)	Edge	M12
		IoT cameras	IoT	Available
		Professional cameras	IoT	M18
		Video-to-IP gateways	IoT	M12
Multimedia Content Distribution	Delivers robust, low-latency media distribution via zero-trust policies and advanced threat	Content PoP	Cloud/Edge	M18
		Video Quality Probe	Cloud/Edge	Available
		Optimizer	Cloud/Edge	M12

	mitigation across multi-domain edge-cloud.	Virtual Networking	Cloud/Edge	M18

3.3.3. Data collection

The following datasets have been identified as part of the pilot:

Table 7. Trial 3 data collection

Dataset	Description/Specifications	Status
Dataset 1	Media Streams	M18
Dataset 2	Media Production Metadata	M12
Dataset 3	Quality Metrics and Telemetry	M18
Dataset 4	User Access & Authentication Logs	M18
Dataset 5	Cybersecurity Event Logs & Alerts	M18
Dataset 6	Infrastructure & Operational Metrics	M12

Table 8. DMP for Trial 3 datasets

Dataset #1	Dataset name: Media Streams
Brief Description	Raw and high-quality contribution streams, low quality previews, and high quality compressed program signals.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	This data is continuously gathered via live feeds, with software-based feeders receiving streams from cameras or gateways. Network monitoring tools (cDNS & AI-FWaaS) oversee traffic flow, while attention is given to privacy considerations—particularly if cameras capture personal or sensitive data—to ensure compliance with regulations such as GDPR.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Media files are saved in standard formats (e.g., MPEG-TS, H.264) using structured file names (e.g., CAMERAID_YYYYMMDD_HHMMSS) that reflect the source device, date, and time of capture. Each file is documented through a concise data dictionary that explains technical parameters like resolution, bitrate, and frame rate.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	Metadata standards such as XMP or container-specific headers store essential file properties (e.g., codec type, capture device ID), supplemented by external records in JSON or XML. Version-controlled documentation

	details each media file's attributes, ensuring consistent references across all system components (e.g., MPE, vCE).
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	Automatic checks verify data integrity on arrival, employing checksums to detect corruption, while real-time monitoring flags anomalies like dropped frames or mismatched bitrates. Contaminated or partial streams can be discarded or re-requested, and uniform formatting is enforced when discrepancies are discovered.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Sharing of raw streams is often restricted due to large file sizes and privacy concerns, with authorized internal stakeholders or external partners accessing data under strict NDAs. Role-based access (ZTNA) and encryption protocols regulate data downloads or real-time viewing, and if partial sharing is necessary, down sampled or anonymized segments are provided.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Long-term retention uses tiered storage, placing frequently accessed files on high-performance media while archiving less active streams to cost-effective solutions. Periodic checks ensure format compatibility, with obsolete codecs converted as needed, and cloud-based backups include regular integrity verifications.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	To protect the data, at-rest encryption and secure transmission protocols (HTTPS, SSL/TLS) are mandatory. Strict access controls and multi-factor authentication prevent unauthorized handling, while data use agreements specify permissible operations on these raw media files, minimizing risks of misuse.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	A designated data steward oversees raw stream compliance with privacy, security, and operational guidelines. A multi-partner committee periodically reviews adherence to this plan, ensuring consistent enforcement and coordination with stakeholders responsible for media production and cybersecurity.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	All processes align with applicable regulations (e.g., GDPR) and organizational data policies, mandating secure infrastructures, staff training, and robust governance procedures. Audits confirm adherence to zero-trust principles and institutional guidelines, verifying that privacy, confidentiality, and integrity requirements are continuously met.

Dataset #2	Dataset name: Media Production Metadata
Brief Description	This includes MPE editing logs (timestamps, edit decisions, etc.), compression parameters (bitrate, format), content enrichment (subtitles, contextual info, tags, ...) and optimizer adjustments (profiles, adaptive bitrate...). So all ancillary information generated and used during content creation and editing.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Metadata is captured automatically whenever users interact with production tools or services (e.g., logging edit decisions in the MPE), and generated by software modules (e.g., vCE's compression settings, EN2's enrichment data). The data collection process remains mindful of privacy obligations, especially if personally identifiable information (PII) emerges in annotations or subtitles.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Files and records are structured around production sessions, project IDs, and version numbers, using naming conventions such as PROJECTID_SESSIONID_TIMESTAMP. These structures are complemented by an index or registry that cross-references files, ensuring that each piece of metadata is traceable back to its corresponding media asset.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	Detailed dictionaries define each metadata field—e.g., compression ratio, subtitle language, clip duration—alongside permissible value ranges. Internal or external (JSON/XML) descriptors capture essential details on how the metadata interacts with other processes, guaranteeing consistency across the production pipeline and between different system components.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	Automated validation checks confirm that metadata adheres to expected formats and semantic rules (e.g., correct timestamp ordering, valid language codes). Any detected inconsistencies—such as missing or invalid fields—trigger alerts for manual review, ensuring production workflows are fed with accurate and consistent data.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Metadata sharing is largely constrained within the production ecosystem but can extend to partners responsible for downstream processing or analytics. Access is regulated by role-based permissions (e.g., producers, editors, machine-learning developers), while external requests require explicit authorization and adherence to established nondisclosure agreements.

Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Long-term storage strategies mirror those of the core media files, linking metadata to the corresponding raw or processed media. Regular backups ensure older metadata versions remain recoverable, and standardized schemas protect against obsolescence by providing an easy path for migration if production software is upgraded.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Encryption at rest and in transit safeguards metadata containing sensitive information like user IDs or potentially identifying details in subtitles. Access control policies enforced through zero-trust network principles (ZTNA) and auditing ensure that only authorized personnel or processes can read or modify the metadata.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	A dedicated metadata steward oversees the integrity and consistency of metadata workflows, with a governance committee reviewing policies for data handling, retention, and format evolution. They coordinate with security teams to maintain alignment with overall project guidelines on data privacy and cybersecurity.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	All metadata handling aligns with institutional and regulatory standards, ensuring that privacy obligations (e.g., GDPR) and organizational policies remain upheld. Regular compliance checks validate that relevant encryption, logging, and consent requirements are consistently applied throughout the media production pipeline.
Dataset #3	Dataset name: Quality Metrics and Telemetry
Brief Description	Performance data, Quality probe output, and network monitoring data.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Metrics are gathered via software agents embedded in production tools (e.g., MPE, vCE, EN2) and distribution components (e.g., Content PoP, Optimizer), as well as network monitors (cDNS & AI-FWaaS). Automated polling and event-based triggers supply continuous updates, respecting privacy rules by excluding sensitive user information unless explicitly permitted.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Measurements are stored in time-series databases, keyed by source component, timestamp, and metric type. Naming conventions (e.g., ComponentID_MetricType_Timestamp) keep records consistent, while indexes support

	efficient retrieval and correlation across the pipeline.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	A data dictionary defines each metric (e.g., CPU_Usage_Percentage, Bitrate_Mbps), outlining valid ranges, units, and calculation methods. Configuration files detail polling intervals, threshold alerts, and the hierarchy of measured components, ensuring cross-system consistency.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	Automated validation checks confirm incoming metrics fall within expected operational ranges or produce an alert if anomalies arise (e.g., sudden spikes or zero values). Periodic manual reviews assess measurement accuracy and calibrate collection agents to maintain reliable performance data.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Real-time dashboards display aggregated metrics to authorized staff (e.g., operations, security teams), providing immediate insight into system health. External data requests, such as for third-party analytics, require approval and may be shared in anonymized or aggregated form to protect proprietary information.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Time-series data with higher granularity is retained for a defined short-term period (e.g., 30 days) to support real-time troubleshooting, then summarized and archived for longer-term trend analysis. Archival systems use mirrored storage to mitigate data loss, and periodic schema updates preserve format compatibility.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	All telemetry traffic traverses encrypted channels to guard against eavesdropping, and role-based policies (ZTNA) restrict who can view or modify telemetry records. Logs and metrics containing potentially sensitive information (e.g., user IDs) undergo pseudonymization or filtering where feasible.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	A performance data steward ensures consistent metric definitions and oversees threshold/alert updates. Regular governance reviews confirm alignment with the overall data management plan, ensuring collected metrics serve operational, security, and compliance needs.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	Telemetry handling aligns with institutional policies, including those governing data retention and privacy (if any personal data is inadvertently captured). Audits verify the

	encryption, access controls, and anonymization protocols used to protect sensitive operational information in accordance with applicable regulations.
Dataset #4	Dataset name: User Access & Authentication Logs
Brief Description	Records of IDs, roles, events, timestamps, and accessed component.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Logs are generated whenever a user or service attempts to authenticate to any component (e.g., MPE, Feeder, Content PoP). Monitoring tools (IPDM-DSS, CASB) collect and correlate these logs in near-real time. All entries capture relevant metadata (e.g., user ID, timestamp, login outcome) while conforming to privacy regulations by restricting sensitive details (e.g., passwords).
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Authentication logs are stored in chronological order within a central logging system or SIEM, often in standardized formats like JSON. They are categorized by service/component (e.g., "Feeder_AuthLog," "MPE_AuthLog") and sub-labeled by event type (success, failure, role escalation), aiding quick retrieval and analysis.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	A data dictionary outlines each field—e.g., user ID, roles, authentication factor used, IP address—and the permissible values or formats for each. A metadata file explains the log structure, retention periods, and how each field maps to system actions or security events.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	Periodic audits and automated validation scripts check for malformed entries (e.g., missing timestamps), inconsistent user IDs, or suspicious anomalies (e.g., repeated failed attempts). Logs deemed invalid are flagged for further review, ensuring accurate threat detection and post-incident analysis.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Internal cybersecurity teams have direct access for monitoring and incident response, while external partners may only receive aggregated or anonymized data, subject to nondisclosure agreements. Role-based access ensures that only authorized users can extract or analyze these logs.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Logs are retained in high-availability storage for a defined period (e.g., 90 days) to support real-time threat detection, then transferred to archival storage for extended retention (often required by compliance). Regular backups and redundant storage mechanisms guard against accidental loss.

Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Encryption at rest and in transit protects log files, and access controls (ZTNA) limit which personnel or automated services can view them. Privacy considerations ensure user identifiers are stored in a way that meets GDPR or similar regulatory requirements, preventing unwarranted disclosure of personal data.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	A dedicated log management owner (often the security officer) ensures consistent log retention policies, coordinates access requests, and organizes routine compliance checks. They work with data stewards and legal teams to maintain alignment with organizational policies and regulations.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	Authentication logs conform to legal guidelines for data protection and incident response record-keeping, ensuring readiness for audits or forensic investigations. Encryption, limited retention, and robust access controls meet the key requirements in standards such as GDPR and relevant cybersecurity frameworks (e.g., NIST).
Dataset #5	Dataset name: Cybersecurity Events Logs & Alerts
Brief Description	Intrusion alerts, flagged activities and logs.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	These logs are generated by intrusion detection/prevention systems (IPDM-DSS), AI-based firewalls (AI-FWaaS), and zero-trust tools (ZTNA, CASB). Each security component forwards event data to a central logging repository, timestamping and categorizing alerts by severity and source to facilitate rapid investigation.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Alerts are stored with metadata fields indicating event ID, timestamp, affected component, threat type, and response actions taken. Files or records follow consistent naming patterns (e.g., SEC_EVENT_YYYYMMDD_HHMMSS.json), and are grouped by system or domain (production, CDN, etc.) to streamline analysis.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	A detailed data dictionary defines each alert type (e.g., "Suspected DDoS," "Port Scan Detected," "Unauthorized Login Attempt"), linking event codes to recommended remediation steps. Records also map to relevant security controls (e.g., IPDM-DSS policies) and detail the response or escalation path.

Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	Automated validation routines check incoming logs for completeness, accuracy, and proper formatting. Security teams review flagged entries or unusual patterns, reconciling data with real-time monitoring dashboards to confirm the validity of each alert.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Immediate access is restricted to cybersecurity personnel and system administrators for prompt incident response, with selective sharing allowed for external security analysts under nondisclosure agreements. If broader research or compliance reporting is required, event data is anonymized or summarized before dissemination.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Critical security events remain accessible in high-availability storage for a defined period (e.g., six months or more), then move to archival solutions that maintain forensic integrity. Regular backups ensure these logs remain intact for post-incident reviews or compliance audits.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Strict encryption mechanisms protect logs at rest and in transit. Access controls aligned with zero-trust principles ensure only authorized users can view or modify alerts. Potentially sensitive data (e.g., IP addresses, partial user identifiers) may be hashed or tokenized to reduce exposure risk.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	A dedicated security officer or event log steward manages this dataset, setting retention schedules and overseeing internal audits. This role collaborates with organizational leadership to maintain consistent log management policies and handle exception requests or unusual threat patterns.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	Cybersecurity event data aligns with regulations and frameworks mandating detailed incident records (e.g., GDPR, NIS Directive, ISO 27001). Comprehensive logging practices ensure readiness for external audits, while organizational protocols define the timeframes and conditions under which event data must be retained or reported.
Dataset #6	Dataset name: Infrastructure & Operational Metrics
Brief Description	Node/instance health data.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Metrics are gathered continuously from monitoring agents running on servers, containers, and virtual networks. These agents

	report system health to a central analytics platform, respecting privacy constraints by collecting primarily technical telemetry rather than user-specific details.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Metrics are stored in a time-series format, indexed by component ID and timestamp (e.g., Server01_CPUUsage_20250410_1030). Aggregations or rollups (hourly, daily) facilitate long-term trend analyses while maintaining fine-grained data for immediate troubleshooting.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	A data dictionary defines each metric's meaning (e.g., CPU usage, network I/O, memory consumption), permissible ranges, and sampling intervals. Configuration files map monitored entities (e.g., server IDs, VM instances) to their respective metric sets for easy reference.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	Automated validation scripts flag anomalies or missing values, comparing observed metrics against predefined baselines. Manual spot checks by operations teams confirm that collected data accurately reflects real-world resource usage, with misconfigured agents corrected promptly.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Operational teams (DevOps, SRE, security) have real-time dashboard access for system monitoring and incident response. Summaries or aggregated data may be shared externally if relevant (e.g., project performance reports), subject to nondisclosure agreements and sanitized to remove proprietary details.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	High-resolution metric data is retained for a limited window (e.g., 14–30 days) to support immediate diagnostics, while aggregated data is archived longer for capacity planning and historical analysis. Periodic backups and checks prevent data loss and ensure ongoing compatibility.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Encryption in transit (TLS/HTTPS) protects metric flows, and at-rest data encryption secures stored telemetry. Access is constrained by role-based policies, ensuring only authorized individuals or services can view or manipulate system metrics.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	A system performance steward oversees metric collection policies, retention schedules, and access permissions. Governance committees periodically review data utility and security measures, aligning them with broader

	organizational standards for data management and cybersecurity.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	Collected metrics comply with any relevant data handling policies, particularly if they could indirectly identify user behaviour. Regular audits verify adherence to institutional guidelines (e.g., ISO 27001), ensuring all telemetry practices support the trial's security and privacy objectives.

3.4. Trial#4: Proactive & Intelligent Protection of Healthcare Critical Infrastructures

3.4.1. Trial description and objective

XGL and MUP are responsible for Trial#4 to evaluate CyberNEMO effectiveness in assuring and granting Proactive & Intelligent Protection of Healthcare Critical Infrastructures through detecting and denying unauthorized access and miss use of the sensitive health records circulating and stored at National Health Information system (NHIS) and National Health Insurance Fund (NHIF) in Bulgaria. The trial aims to evaluate CyberNEMO solutions to verify their ability to identify, respond to and mitigate the cyber threats to the health data management Medical University - Plovdiv Hospital (MUP) and medical centres are the infrastructure support for Trial#4 which are presenting the entry and end points for health data records both for pre-hospital and hospital medical healthcare. The password protected access and gamma-consult information system used in medical facilities for processing the data prior to its inclusion into NHIS and NHIFS are protecting the health records from malicious interventions. The goal is to secure the cloud continuum (consisting of IoT, edge devices, and central cloud) by applying CyberNEMO ZTNA, which requires authentication and authorization for every process accessing NHIS and NHIF.

Use case applications:

- Detecting and Mitigating Authentication and Authorization Vulnerabilities & Threats: Protecting patients' health records, as well as medical centres' and hospitals' data, is crucial to preventing unauthorized access. Strong authentication and authorization measures prevent cyber threats targeting healthcare infrastructures and individual patients' healthcare records. Risks include exfiltration and or infiltration of sensitive health data, manipulation of data related to the medical logistic and medical devices, consumables and pharmaceuticals purchase along with financial reports and invoices. The aim of this use case is to ensure the protection of healthcare data and the security of HIS and NHIF, safeguarding sensitive information from breaches
- Supply Chain Collaboration for Increased Resilience: Security collaboration such as sharing adversarial tactics, remedial best practices, and zero-day vulnerabilities across healthcare supply chains can strengthen the protection of information exchange and storage infrastructure against cyber threats. This ensures reliable healthcare data management at all levels, from individual and pre-hospital care to hospital, local, regional, and national systems. By analysing and mitigating the cyberattacks on Healthcare Critical Information Infrastructures, the operability and functionality of the entire healthcare system can be improved. This collaborative approach also helps reduce response times to cyber threats, enhancing overall resilience and security.

3.4.2. Required Equipment/Assets

Table 9 below shows a detailed description of the required equipment for the scenarios, along with its type and status (whether it is available or still to be procured). The mentioned equipment/assets are available and operational now. For validation of the CyberNEMO solution technologies and how they could be tested with the

operational systems this will be clarified at later stage, depending on the requirements of the CyberNEMO and on the HHIS providers.

Table 9. Detailed description of equipment/assets in Trial 4.

Scenario	Scenario description	Equipment/Asset	Type	Status
Preventing unauthorized access	This scenario involves a hospital's cloud and edge devices (such as smartphones and scanners used to access patient records), where access to the Hospital Information System (HIS) can be granted without a doctor's approval. It also includes instances of unauthorized access to data from other departments.	HIS	Central Cloud	Available
		Local hospital Nodes	Edge device	Available
		Smartphones/tablets	IoT Device	Available
Protection of central and local clouds	This scenario involves a sophisticated cyberattack targeting a hospital's systems, exploiting phishing and network interception to gain unauthorized access. The attacker sends a phishing email impersonating the Ministry of Health, tricking a staff member into providing login credentials, which grants access to the hospital's secure WAN and Hospital Information System (HIS). Once inside the system, the attacker alters patient records in the Electronic Health Records (EHR) database to include false or manipulated medical information. For personal financial gain, they exploit the compromised data to file fraudulent claims with the National Health Insurance Fund (NHIF) for medical services that were never provided or attributed to fake patients. This not only results in financial losses for the hospital but also damages its reputation, undermining trust in its ability to manage sensitive healthcare data securely.	HIS and NHIF	Central Cloud	Available
		Local hospital Nodes	Edge device	Available
		Smartphones/tablets	IoT Device	Available
Cyber Threats to the networking equipment	This scenario involves an internal/external attacker gaining unauthorized access to the hospital's local network through packet monitoring and network cracking techniques, targeting data communication between the	Servers, routers, switches, firewalls, and other networking equipment	Edge	Available

	Hospital Management Software system, the Hospital Information System (HIS), and the National Health Insurance Fund (NHIF). By intercepting and analyzing network traffic, the attacker exploits vulnerabilities to manipulate data in transit, allowing them to update records maliciously from within the network. These modifications can include altering patient records, falsifying billing information, or injecting incorrect data into NHIF claims to facilitate financial fraud. This intrusion undermines the integrity of hospital systems, disrupts operational workflows, and compromises both the hospital's reputation and its ability to securely manage critical healthcare data.			
Operational healthcare supply chain	This scenario involves communicating adversarial practices, remedial best practices for known threats (e.g., ransomware, medical devices attacks), as well as information about potential zero-day vulnerabilities in the healthcare supply chain. This involves securely communicating this information from Central cloud (Ministry) to the hospital (HIS and NHIF) to practitioners and staff (smartphones and scanner of medical documents)	HIS and NHIF	Central cloud	Available
		local clouds at the hospital	Edge node	Available
		smart phones and tablets	Mobile/IoT Devices	Available

3.4.3. Data collection

The data details for this pilot will be available by M10. This timeline reflects the standard data preparation process involving an external partner, whose contributions are planned and ongoing. The data is being generated and organized according to the project's schedule, and no issues are currently anticipated.

3.5. Trial#5: Smart Farming and FinTech/Logistics Supply Chain

3.5.1. Trial description and objective

The trial combines smart farming and the logistics supply chain. It combines smart agricultural data, retrieved through multiple types of ground micro-climate/soil/leaf information stations, and logistics-related information to ensure traceability and auditability ensuring end-to-end cybersecurity.

The trial includes two main use cases to be demonstrated, validating CyberNEMO:

Auditing mechanism to safeguard accountability based on evidence and not only trust.
The 1st use case of the pilot develops an audit service to identify and verify points of failure affecting product quality and to clarify disputes on specific aspects of cultivation such as bio certification. The use case monitors the growing conditions of an organic Olive trees' plantation (using sensors, drones and agri-robots with multispectral cameras), the collection, milling of extra virgin oil with the highest quality standards, bottling, storage and transportation conditions to the selling point.
Through the use case the stakeholders and consumers will be able to trace the origins of products, the specific stages they've traversed, identifying at the same time any potential products that might be affected in instances of quality-related issues.
The trial application increases transparency and trust in auditing processes, especially in cases of a) safety and quality issues and b) disputes on bio certification (e.g. validating the usage of insecticides).

Secure and traceable smart F2F trading.
The 2nd use case of the pilot is focused on applications related to the supply chain, executed in an automated, trusted, traceable, and decentralized way. Such applications involve supply chain stakeholders and trading aspects such as agreements, financial aspects and transaction verifications. The use case applies cybersecurity and privacy protecting federation of (otherwise siloed) IoT platforms, establishing a distributed and immutable data management case to support traceability and quality control of transported items.
The use case enhances trading features within the smart agriculture and supply chain in an automated, trusted, and decentralised way allowing for security, transparency, and traceability, increasing competition among the supply chain actors (buyers and sellers). The actors agree on the terms of the trade, perform payments, and verify that the trade has been carried out successfully with as little user interaction as possible. The case will also involve auditability for potential dispute resolutions.
Through CyberNEMO infrastructure and resources, the pilot application will contribute to transparency and security in supply chain trading processes, supporting a) transparency and traceability in data exchange in supply chain trade-related applications and b) automation and high performance of applications executed in federated environments.

3.5.2. Required Equipment/Assets

Table 10 below shows a detailed description of the required equipment for the scenarios, along with its type and status (whether it is available or still to be procured).

Table 10. Detailed description of equipment/assets in Trial 5.

Scenario	Scenario description	Equipment/Asset	Type	Status
Scenario 1 Irregular traffic by compromised IoT device (related to UC1)	Compromised IoT Device sends irregular/non-typical/erroneous data. The compromise may be due to malicious firmware update. The node should be quarantined.	SynField node	IoT node collecting sensor measurements on environmental, crop and soil conditions	Available
		Edge server	It performs data persistence and edge operations	Available
		Cloud server	It hosts server – side components	Available
IoT node (with actuator	The IoT node with actuation capabilities	Agricultural drone and agri-robot	IoT nodes	M12

capabilities) malfunctioning (related to UC1)	(e.g. drone) is malfunctioning, due to man-in-the-middle attack and the battery is quickly diminishing. Task off-loading towards the edge and/or cloud has to take place.	Edge server	It performs data persistence (e.g. images) and edge operations	M12
		Cloud server	It hosts server – side components (such as ML training)	M12
Low performance in traceability, quality control and trading applications (related to UC2)	Traceability and quality control applications using data from different IoT platforms and involving data registration in DLTs experience low performance and delays (e.g. as a result of DoS attack)	IoT nodes	IoT nodes (SynField) offer data	M12
		Edge server	It performs data persistence as well as quality control and trading applications	M12
		Cloud server	The cloud server accommodates the quality control and trading services as well as the DLT nodes	M15
Attacks on FML mechanism (related to scenario 2)	Attacks on FML mechanisms which are based on data/logs of far & near edge nodes and applications involved in smart agriculture and / or supply chain applications residing in different domains	IoT nodes	IoT nodes (SynField) offer data for ML and FML.	Available
		Edge server	The edge server performs data persistence and hosts FML components (client part)	M12
		Cloud server	The cloud server accommodates FML components (aggregator part) and the DLT nodes	M12

3.5.3. Data collection

The following datasets have been identified as part of the use case:

Table 11. Trial 5 data collection

Dataset	Description/Specifications	Status
Dataset 1 – Smart farming / supply chain application-level measurements	Data retrieved from the applications in the context of traceability, auditing, and trading.	Not available / under design

Dataset 2 – Platform / Network measurements	Data retrieved from the platform / network as used by the pilot applications	Not available / under design

Table 12. DMP for Trial 5 datasets

Dataset #1	Dataset name: Smart farming / supply chain application-level measurements
Brief Description	Smart Farming / Supply Chain Application – level measurements. This dataset will include time-specific measurement and data retrieved by the nodes (including IoT) participating in the pilot as well as the application logs as executed in the edge / cloud.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The dataset will be collected through the logs of the applications and IoT nodes participating in the pilot use cases
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	We foresee that the data collection will be timeseries-based. The data model of the participating IoT infrastructures will be used (e.g. SynField). The naming and metadata rules defined in the DMP will be followed.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	We consider using metadata from the DCAT standard to describe the dataset
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	We will apply 1st level data quality checks (on missing values, repetitions, inconsistencies) and 2nd level rule-based validation
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The dataset will be anonymized and will be distributed initially to the consortium. Provision to repository such as Zenodo will also be considered.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The pilot technical leader will take care of the dataset during the lifetime of the project. External repository will be considered after the end of the project.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The dataset will be anonymized. The exact technique (removal, aggregation, encryption etc) will be decided at a later phase considering the requirements of the applications (including the AI/FML considerations).
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	In terms of data governance, the key role will be owner of the application-level data (i.e. the infrastructure – Aghia Sofia owner). The curation will take place by the technical leader of the pilot.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data management plan will comply with the GDPR (and will consider the Data Act and the Data Governance Act)

Dataset #2	Dataset name: Platform / Network measurements
Brief Description	Smart Farming / Supply Chain network – level measurements. This dataset includes data retrieved from infrastructure networking and computing nodes, typically with the support of the owner of the network infrastructure and the application provider.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The dataset will be collected through the usage of network monitoring tools (OTE is employing Prometheus), and the monitoring of Kubernetes environment. Security – related tools (Firewall, IDS provided by the project) will also contribute to this dataset providing information on relevant events.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	We foresee that the data collection will be timeseries-based. In principle we will follow the data model adapted by the monitoring / reporting tool (e.g. Prometheus, Suricata), possibly employing open standards such as (openmetrics).
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	We consider using metadata from the DCAT standard to describe the dataset
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	We will apply 1st level data quality checks (on missing values, repetitions, inconsistencies) and 2nd level rule-based validation to identify and remove inconsistencies
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The dataset will be anonymized and will be distributed at first phase to the consortium. The provision of the dataset to repository such as Zenodo will also be considered.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The pilot technical leader will take care of the dataset during the lifetime of the project. External repository will be considered after the end of the project.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The dataset will be anonymized. The exact technique (removal, aggregation, encryption etc.) will be decided at a later phase considering the requirements of the applications (including the AI/FML considerations).
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	In terms of data governance, the key role will be the owner of the data, i.e. the network, platform and application providers. The curation will take place by the technical leader of the pilot.
Compliance: how the data management plan aligns with relevant laws, regulations, and	The data management plan will comply with the GDPR (and will consider the Data Act and the Data Governance Act)

institutional policies, including data privacy and security regulations.	

3.6. Trial#6: Pan-European SAAM multi-Living Labs Federation

3.6.1. Trial description

Trial 6 focuses on validating the federation of CyberNEMO pilots across diverse environments and application domains. It will demonstrate seamless access to heterogeneous IoT ecosystems (ROS, Android, iOS, Linux) and test secure, anonymized data sharing across five countries: France, Italy, Spain, Greece, and Bulgaria. The trial objectives are:

- Enable cross-platform federation for secure, trusted service deployment.
- Validate the SAAM framework and governance models at a pan-European scale.
- Aggregate AI models using Zero Trust Federated Machine Learning (ZT-FML) and Transfer Learning.
- Test CI federation for auditing, certification, and accreditation of micro services.

This trial will showcase CyberNEMO ability to enable scalable, federated services while ensuring data sovereignty and cyber secure operations, validated by global stakeholders like DNV, ENG, TSG, SIEM, and SPACE.

4. Conclusion and Future Steps

This deliverable introduces the initial DMP for the CyberNEMO Project, detailing the processes for managing, storing, and sharing datasets across trials. Future iterations will assign Data Manager Officers to each pilot to oversee data collection, quality management, and site-specific DMP implementation.
The DMP is a dynamic document, updated regularly to reflect project progress and ensure compliance with ethical and legal frameworks, particularly for trials involving personal data.
The next steps involve assigning Data Manager Officers for each pilot to ensure effective oversight. Additionally, open-source and open-data policies will be developed to facilitate ethical data sharing. The next version (M18) will incorporate updates to trial-specific methodologies and cross-site procedures, ensuring consistency across implementations. Initial results will be integrated to align with pan-European data governance frameworks. Finally, the remaining pilot setups will be completed, structured data generation and sharing will commence, and representative dataset samples will be defined to support validation in the next phase.

5. References

[1] European Parliament and the Council, "General Data Protection Regulation (GDPR) (EU) 2016/679.," 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

[2] European Union, "General Data Protection Regulation (GDPR) (EU) 2016/679.," 2016. [Online]. Available: https://gdpr-info.eu/.

[3] European Commission, SEDIA, "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020," 2017. [Online]. Available: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm.

[4] European Commission, 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[5] European Commission, "Horizon Europe Programme Guide," 2021. [Online]. Available: https://ec.europa.eu/programmes/horizon2020/.

[6] European Commission, 05 05 2021. [Online]. Available: https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf.

[7] M. D. e. a. Wilkinson, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data,* vol. 3, no. 160018, p. no. 160018, 2016.

[8] IoT-NGIN European Project, GA No.957246, [Online]. Available: https://iot-ngin.eu.

[9] NEMO Horizon EU-funded project, GA No.101070118, [Online]. Available: https://meta-os.eu/.

[10] European Commission, "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).," Regulation (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2016.

[11] European Parliament and the Council, "Artificial Intelligence Act (EU Regulation 2024/1689)," 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.