Net-Zero self-adaptive activation of distributed self-resilient augmented services

D1.2 Quality Assurance, Risk Assessment, Data Management Plan, Ethics & Regulatory issues

Lead beneficiary	ELTE	**Lead author**	Sándor Laki
Reviewers	Edgardo Montes de Oca (MONT), Konstantinos Lampropoulos (P-NET), Virgilios Passas (CERTH), Antonios Lalas (CERTH)		
Type	R	**Dissemination**	PU
Document version	V1.0	**Due date**	30/06/2024

Co-funded by the European Union

6GSNS

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

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

UK Research and Innovation

Project information

Project title	Net-Zero self-adaptive activation of distributed self-resilient augmented services
Project acronym	NATWORK
Grant Agreement No	101139285
Type of action	HORIZON JU Research and Innovation Actions
Call	HORIZON-JU-SNS-2023
Topic	HORIZON-JU-SNS-2023-STREAM-B-01-04
Start date	01/01/2024
Duration	36 months

Document information

Associated WP	WP1
Associated task(s)	T1.1, T1.2, T1.3, T1.4
Main Author(s)	Sándor Laki (ELTE)
Author(s)	Sándor Laki (ELTE), Eryk Schiller (UZH), Nasim Nezhadsistani (UZH), and all partners
Reviewers	Edgardo Montes de Oca (MI), Konstantinos Lampropoulos (P-NET), Virgilios Passas (CERTH), Antonios Lalas (CERTH)
Type	R – Document, report
Dissemination level	PU - Public
Due date	M6 (30/06/2024)
Submission date	30/06/2024

Document version history

Version	Date	Changes	Contributor (s)
v0.1	01/05/2024	First proposal for the table of content	Sándor Laki (ELTE)
v0.2	14/06/2024	Contribution in questionnaires	All partners
v0.3	17/06/2024	Initial version ready for internal review	Sándor Laki (ELTE)
v0.4	27/06/2024	Revised version, sent to the coordinator	Sándor Laki (ELTE)
v0.9	28/06/2024	Final review and refinements	Antonios Lalas (CERTH), Anastasios Drosou (CERTH)
v1.0	30/06/2024	Final version for submission	Antonios Lalas (CERTH)

Disclaimer

Copyright message

Contents

List of acronyms and abbreviations

Abbreviation	Description
AGA	Annotated Grant Agreement
AI	Artificial Intelligence
DMP	Data Management Plan
DoA	Description of the Action
DOI	Digital Object Identifier
DPO	Data Protection Officer
FAIR	Findable, Accessible, Interoperable, and Re-usable
GA	Grant Agreement
GDPR	General Data Protections Regulation
IM	Innovation Manager
KPIs	Key Performance Indicators
ML	Machine Learning
PC	Project Coordinator
QA	Quality Assurance
QARM	Quality Assurance & Risk Manager
QCB	Quality Control Board
QMP	Quality Management Plan
TM	Technical Manager
WP	Work Package

List of figures

List of tables

Executive summary

This document includes the Quality, Risk and Data Management Plans and the Ethics Manual for the NATWORK project. The purpose of this document is to establish project documentation standards and a full set of project procedures to meet the highest quality level of the project outcomes.

The document first describes the Quality Management Plan to ensure that quality standards are met throughout the project's duration. The introduced quality management life cycle aims to ensure that the project will meet its objectives and delivers high-quality outputs. Adhering to ISO 9001 standards, the Quality Assurance (QA) activities are designed to ensure consistency, reliability, and efficiency throughout the project life cycle.

The document also describes the Risk Assessment Plan which is needed for identifying, analyzing, and evaluating potential risks that could affect the project's success in time. The NATWORK project will follow a comprehensive risk assessment process enabling project managers to pinpoint potential risks early, understand their likelihood and potential impact, and develop effective strategies to mitigate or avoid them.

The document also describes the data management life cycle for the data to be collected, processed and/or generated by NATWORK, to make research data findable, accessible, interoperable, and re-usable (FAIR). The NATWORK project applies the FAIR data handling principle and prefers publishing the project results openly. This deliverable also includes a questionnaire on various aspects of data generation, usage and handling filled by all the partners.

Finally, the NATWORK project is committed to adhering to ethical principles and legal frameworks concerning data handling, sharing and the use of Artificial Intelligence (AI) throughout its project life cycle.

1. Introduction

1.1. Purpose and structure of the document

According to the Annotated Grant Agreement (AGA) [1], this deliverable includes a) the Quality Management Plan of the project, b) the Risk Assessment Plan of the project, c) the Data Management Plan abiding to FAIR principles, and d) the Ethics manual of the project. In all sections, we consider the regulatory aspects and the imposed restrictions.

The Quality Management Plan (QMP) describes the objectives, scope, and methodology for quality management in the project. It serves as a guide for all QA activities, ensuring that every deliverable meets the predefined quality standards.

The Risk Assessment Plan provides a comprehensive risk assessment process that enables project managers to pinpoint potential risks early, understand their likelihood and potential impact, and develop effective strategies to mitigate or avoid them.

The Data Management Plan (DMP) describes the involved data and how findable, accessible, interoperable, and re-usable (FAIR) data is provided. Then, resources (personnel and infrastructure) are allocated for data management and finally ethical aspects are discussed.

Finally, the Ethics Manual presents the relevant ethical principles, legal frameworks, and guidelines for trustworthy AI, and describes that we aim to conduct our research responsibly and ethically.

The document is structured as follows:

- In **Section 1 – Introduction**, general information about the document, the intended audience and interrelations;
- In **Section 2 – Quality Management Plan**, guidelines for ensuring high quality project results are provided;
- In **Section 3 – Risk Assessment Plan**, the identified implementation risks and relevant mitigation measures are described;
- In **Section 4 – Data Management Plan**, the plan for providing FAIR data and results is presented;
- In **Section 5 – Ethics Manual**, ethical considerations, in particular related to AI development, are described;
- In **Section 6 – Other issues**, other relevant issues are discussed;
- In **Section 7 – Conclusions**, concluding remarks are provided.

1.2. Intended audience

The NATWORK Project's Quality Assurance, Risk Assessment, Data Management Plan and Ethics Manual is devised for public use in the context of project management and dissemination/ communication activities of the NATWORK consortium, comprising members, project partners, and affiliated stakeholders. This document mainly focuses on guidelines for quality assurance, risk management and mitigation, data management and ethics considerations, thereby serving as a referential tool throughout the project's lifespan.

1.3. Interrelations

The NATWORK consortium integrates a multidisciplinary spectrum of competencies and resources from academia, industry, and research sectors, focusing on user-centric service development, robust economic and business models, cutting-edge cybersecurity, seamless interoperability, and comprehensive on-demand services. The project integrates a collaboration of fourteen partners from ten EU member states and associated countries (UK and CH), ensuring a broad representation for addressing security requirements of emerging 6G Smart Networks and Services in Europe and beyond.

NATWORK is categorized as a "Research Innovation Action - RIA" project and is methodically segmented into 7 WPs, further subdivided into tasks. With partners contributing to multiple activities across various WPs, the structure ensures clarity in responsibilities and optimizes communication amongst the consortium's partners, boards, and committees. The interrelation framework within NATWORK offers smooth operation and collaborative innovation across the consortium, ensuring the interconnection of the diverse expertise from the various entities (i.e., Research Institutes, Universities, SMEs, and Large industries) enabling scientific, technological, and security advancements in the realm of 6G. The Quality Assurance, Risk Assessment, Data Management Plan and Ethics Manual addresses key management activities of the NATWORK project workplan.

2. Quality Management Plan

The Quality Management Plan (QMP) is an essential component of the NATWORK project, ensuring that quality standards are met throughout the project's duration. The quality assurance (QA) process is crucial for ensuring that the NATWORK project meets its objectives and delivers high-quality outputs. Adhering to ISO 9001 standards, the Quality Assurance (QA) activities are designed to ensure consistency, reliability, and efficiency throughout the project life cycle.

2.1. Adherence to ISO 9001 Standards in Quality Assurance Activities

Adhering to ISO 9001 standards, the Quality Assurance (QA) activities within the NATWORK project are designed to ensure consistency, reliability, and efficiency throughout the project lifecycle. This section outlines how these standards are integrated into our QA processes to achieve high-quality outcomes.

2.1.1. Consistency

ISO 9001 emphasizes the need for consistent processes and outcomes, which is crucial for maintaining the project's integrity. All QA activities will be documented in detail, providing clear guidelines for each project stage. Furthermore, NATWORK will use standardized templates and forms for documentation and reporting, ensuring uniformity across all deliverables. Regular reviews and assessments will be conducted at predefined intervals to ensure all processes are followed consistently. Furthermore, comprehensive checklists will be used during reviews to verify that all necessary steps are completed in accordance with the standards. Continuous training programs will be conducted to ensure all team members are familiar with ISO 9001 requirements and the importance of consistent application. New team members will undergo QA induction sessions to familiarize themselves with the standardized procedures and expectations.

2.1.2. Reliability

Reliability is a crucial principle of ISO 9001, ensuring that the project outputs are dependable and meet the required standards. Clear objectives are fixed at the beginning of the project, which is aligned with ISO 9001 principles. Potential risks are identified in the project proposal and initial mitigation strategies are considered. Furthermore, the risk assessment plan is

provided to identify risks that can occur at any point in the project execution. Furthermore, NATWORK will implement rigorous testing protocols to verify that deliverables meet the defined quality standards. Moreover, NATWORK will maintain detailed audit trails to track the history of changes and ensure the traceability and reliability of data. Regarding non-conformance handling, NATWORK will establish procedures for promptly identifying and addressing non-conformities to prevent a recurrence. Using feedback from reviews and audits, NATWORK will continuously improve processes and enhance reliability.

2.1.3. Efficiency

Efficiency is critical for ensuring that project resources are used effectively, and deliverables are produced on time without compromising quality. NATWORK will adopt lean processes to eliminate waste and improve the efficiency of QA activities. Moreover, NATWORK will utilize automated tools and software for tracking, reporting, and managing QA processes to reduce manual effort and errors. Furthermore, NATWORK will plan resources effectively to ensure that QA activities are conducted efficiently without overburdening the team. Furthermore, NATWORK will leverage the skills and expertise of team members appropriately to maximize productivity and efficiency. Furthermore, NATWORK will continuously monitor performance metrics like the state of KPIs defined in the proposal to ensure that necessary adjustments can be made if needed.

2.2. Objectives, Scope, and Methodology of the QMP

The QMP outlines the objectives, scope, and methodology for quality management in the project. It serves as a guide for all QA activities, ensuring that every deliverable meets the predefined quality standards.

2.2.1. QMP Objectives

The primary objectives of the QMP are to define quality standards, ensure compliance, enhance project efficiency, promote continuous improvement, and facilitate stakeholder satisfaction. Defining quality standards involves establishing clear and measurable criteria for all project deliverables aligned with ISO 9001 standards and other relevant benchmarks. Ensuring compliance requires that all project activities and deliverables meet regulatory and internal policy requirements. Enhancing project efficiency involves implementing effective quality management processes to optimize resource use, minimize waste, and reduce rework. Promoting continuous improvement focuses on fostering a culture of ongoing evaluation and

updating quality management practices, including implementing corrective and preventive actions based on feedback and audits. Facilitating stakeholder satisfaction means understanding and meeting stakeholder needs and expectations, enhancing their confidence in the project's quality management processes and outcomes.

2.2.2. QMP Scope

The scope of the QMP covers all aspects of the project that affect quality. This includes all phases of the project lifecycle, from planning and design to implementation and delivery, encompassing activities related to project management, technical development, and support functions. It also includes all project outputs, such as reports, software, documentation, and other products, as well as interim deliverables, milestones, and final outputs. Additionally, the scope addresses all internal and external stakeholders, including project partners, end-users, and regulatory bodies, and outlines the stakeholder engagement and communication processes. Finally, it encompasses the quality management processes, including procedures for quality planning, control, assurance, and improvement, along with methods for risk management, issue resolution, and performance measurement.

2.2.3. QMP Methodology

The methodology for quality management in the NATWORK project ensures a systematic and structured approach to maintaining high standards. Quality planning involves:

- Identifying quality requirements for each deliverable based on stakeholder needs and project objectives.
- Defining specific, measurable quality objectives and metrics to monitor progress and performance.
- Developing a comprehensive strategy for meeting these quality requirements throughout the project.

Quality control entails implementing inspection and testing procedures to verify that deliverables meet quality standards, establishing review and approval processes involving multiple stakeholders to ensure a thorough evaluation, and defining procedures for managing non-conformities and defects. Quality assurance consists in conducting regular audits and reviews to assess compliance with quality standards and identify improvement areas, continuously monitoring quality performance using predefined metrics and KPIs, and implementing corrective and preventive actions to address identified issues and prevent recurrence. Continuous improvement focuses on establishing mechanisms for collecting and

analyzing feedback from stakeholders and project team members, regularly evaluating quality management processes to identify opportunities for improvement and promoting knowledge sharing and best practices within the project team to enhance overall quality.

2.3. Quality Control Board

The Quality Control Board (QCB) for the NATWORK project will consist of the following people:

- Anastasios Drosou – Project Coordinator (PC)
- Antonios Lalas – Deputy Project Coordinator (PC) and Technical Manager (TM)
- Edgardo Montes de Oca – Innovation Manager (IM)
- Eryk Schiller – Quality Assurance & Risk Manager (QARM)

Additionally, other internal members from the project partners, who are senior researchers with extensive expertise, will be appointed for reviewing specific deliverables and reports. These senior researchers will not include the authors of the specific deliverables.

2.4. NATWORK quality planning, control, and improvement

NATWORK's QMP encompasses quality planning, control, and improvement. Non-conformities are promptly addressed with corrective actions to prevent recurrence, ensuring continuous adherence to quality standards. Deliverables undergo a rigorous review process involving multiple stakeholders. Each deliverable will be reviewed by at least 2 internal members (same WP) and by the PC, the TM and the QARM, 21 days prior to its submission date. The reviewers have 5 working days to complete their reviews and send them to the editor of the document. Criteria for evaluation include compliance with project requirements, accuracy, and completeness. The QCB schedules reviews to align with project milestones and deadlines. Quality documentation is maintained, including all records of QA activities, findings, and corrective actions. Standard templates for reports and forms ensure uniformity and ease of tracking. Regular training sessions are conducted to ensure all project staff are well-versed in QA standards and procedures. Competence requirements for QA roles are clearly defined and monitored. Continuous improvement is a core principle of our QA strategy. Regular feedback loops, lessons learned, and periodic updates to the QA plan ensure that NATWORK adapts and enhances our processes throughout the project.

The chain of revisions and quality assurance is depicted in Figure 1, which outlines its process for all project outputs. To maintain quality standards, each deliverable is reviewed by the respective WP leader. After the first validation, the QCB is responsible for finalizing the revision.

If the work's quality does not meet the requirements, feedback is redirected to WP leaders with the respective comments to create an improved version in a new iteration. This process ensures that all outputs meet the required quality standards before they are approved and implemented.

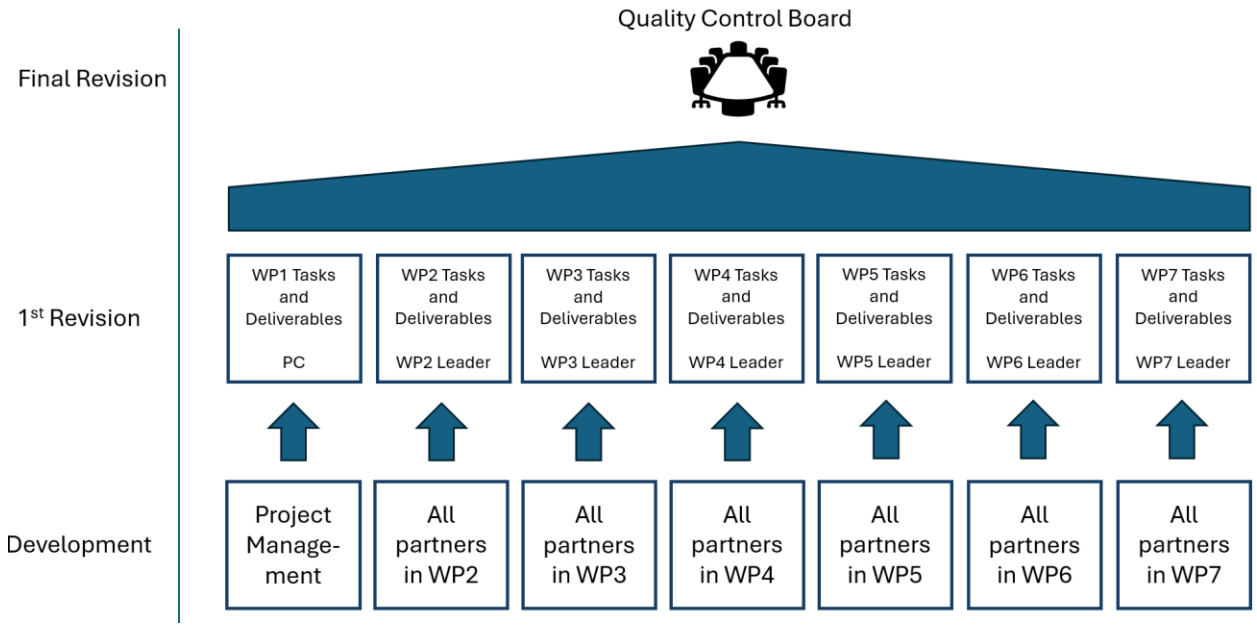

Figure 1: Organization and workflow of the revision process. Each assigned person is responsible for assuring the first validation of the quality of tasks and documents before the QCB.

3. Risk Assessment Plan

3.1. Introduction

Risk assessment is a critical component of project management, essential for identifying, analyzing, and evaluating potential risks that could affect the project's success. In the NATWORK project, a comprehensive risk assessment process enables project managers to pinpoint potential risks early, understand their likelihood and potential impact, and develop effective strategies to mitigate or avoid them.

3.2. Comprehensive Risk Assessment

A detailed risk assessment is presented in Table 1, which outlines the risks identified during the proposal phase along with corresponding mitigation strategies. This assessment is not static; it is designed to evolve throughout the project's duration, accommodating new risks that may emerge as the project progresses.

As the project develops, specific risks that were not apparent initially may arise. The risk assessment process must be adaptive, allowing for continuous monitoring and reassessment of risks. This dynamic approach ensures that emerging risks are promptly identified and addressed.

The evolving regulatory landscape of the European Union related to the project's domain adds an additional layer of complexity. Since the EU's rules and guidelines are constantly changing and can directly or indirectly impact the project, the consortium is committed to closely and continuously monitoring these developments. This proactive stance allows the project to anticipate and adapt to regulatory changes, thereby minimizing potential disruptions.

Table 1: Critical risk for implementation in NATWORK.

Description of risk	Likeli hood	Severity	WP(s)	Proposed risk-mitigation measures
A. General Risks (Management & Dissemination Risks)				
Underperforming partners; low quality of work/ deliverables; systematic delays,	Med	High	ALL	Such issues will be clarified on the Quality Plan and CA. Proper internal peer review procedures will be in place, to ensure quality of the deliverables and their preparation in a timely manner. Regular WP & technical meetings will

etc.				be held to ensure that activities are streamlined and that lessons learnt are shared.
Technical/ administrative disagreement and cooperation problems among partners	Med	Med	WP1	Continuous communication between all the partners. / The PC, STM and QAM will work on problem solving during the project. If necessary, the Plenary Board will decide the right solution according to the CA. The Quality Plan will define the communication procedures and the use of communication tools will be encouraged. The PC is the responsible of solving communication problems, establishing communication flows and methods and calling to bilateral meetings if necessary.
Disputes over ownership of IPR amongst consortium partners	Low	Med	WP1	Standard IPR and access rights clauses will be included in the CA, will be signed before work starts in order to avoid future disputes. The consortium has already discussed these aspects during the proposal phase for avoiding such problems.
Limited or inadequate resources to manage the project complexity	High	Med	WP1	The NATWORK consortium members have long experience in large-scale pilots and large technology-driven innovation projects as well as in the implementation of large and complex systems, thus the possibilities of such problems compromising the project are relatively low. However, the limited financial resources and extended project duration may increase the respective risk. In response to this, the PC will closely and quarterly monitor resource allocation, while the STM will focus on the efficient use of the available personnel resources throughout the project duration. Finally, pilots planning and preparation will examine in depth potential non-anticipated costs and recommend alternative solutions at early stage.
Partner leaves Consortium	Low	Med	WP1	The key expertise areas will be covered by more than one partner, thus ensuring that the project will not suffer until replacement is found or until the partner returns.

Failure to properly disseminate the results of the project towards target stakeholders	Med	Med	WP7	Careful definition and implementation of proper dissemination, communication, and exploitation strategies. Presence in the consortium of partners that are in a very good position to approach relevant European actors.
B. Technological Risks				
Inappropriate definition of reference pilot scenarios	Low	Med	WP6	The expertise of industrial partners, the collaboration between all partners, and the interactions with the Advisory Board will mitigate this risk. Also, the availability of four different pilots targeting different domains will mitigate this risk.
Incomplete requirements analysis and/or prioritisation of requirements.	Low	Med	WP2	Early and systematic identification of the stakeholders and use cases. Internal validation of requirements. Cross- checking of relevance and generality of requirements through project's IAB. Incremental approach with two evaluation cycles provides possibility to rectify risk. Close collaboration between industrial and research partners will minimise the risk of erroneously disregarding important requirements and/or including unrealistic or too ambitious requirements.
Difficulties to implement new functionalities in demos due to lack of flexibility of hardware	Med	High	WP3, WP4, WP5, WP6	Demo setup will be started early enough to assess the risks. Manufactures and their development units are strongly committed to the testbeds and will ensure that the required hardware is provided.
WASMs runtime integrity is unsolvable (without TEE)	Med	Med	WP3	NATWORK will analyse the different possibilities to overcome the SoTA's known weaknesses, without a guarantee of success. The multidisciplinary consortium will devise alternative approaches if required.
Control flow time and frequency metadata extraction or exploitation cannot be done.	Low	Med	WP3	The metadata enriches the monitoring of the software execution. Their extraction, storage and exploitation shall all be defined with care to prevent overload and data scaling. Moreover, the correct definition of the time and frequency series (e.g., where collected, time reference used) shall be also defined with care as it bears the trust and performance impact, crucial for the adoption of the solution.

AI algorithms not leading to improved operational schemes	Med	High	WP3, WP4, WP5	Several algorithms will be employed within WP3-5, and the best will undergo iterative optimisation. Nevertheless, the optimised/standard services will be used as default in case of suboptimal algorithmic performance.
Data analysis algorithms are too resource intensive to be run on the deployed hardware	Med	Med	WP3, WP4, WP5	Hardware is selected under consultancy of the data analysis experts who implement the solutions. The components will be developed and tested on the hardware iteratively. Available HPC infrastructures will be utilised and enhanced capabilities will be employed if required.
Integration of heterogeneous systems fails	Med	Med	WP6	The design and implementation of components will be strictly decoupled from all tool-specific details. Interfaces will be compatible with the existing standards.
C. User-Related Risks				
Unwilling partners to cooperate with the social, ethical and legal requirements.	Low	High	WP1, WP2, WP6, WP7	Compliance with the impact assessment framework will be continuously monitored and should any issues arise the EAB and the PC will take corrective actions through the management structures and procedures defined in Section 3.2.
Limited acceptance by the end-users and relevant stakeholders	High	Med	WP2, WP6, WP7	Well defined user requirements definition and baseline, along with cost-benefit validation of the solution. DCM will follow up and monitor the user requirements accomplishment to ensure methodological vigilance. The evaluation of the solution during pilots will assess user/ stakeholder acceptance and identify room for improvements. Also, various dissemination activities will be carried out to raise the awareness and increase the interest into the project results. The consortium strong links with groups of stakeholders, which already indicated their interest by LoS.
D. Legal/Organizational Risks				
Legal/ Institutional restrictions imposed in the execution of the trials	High	Med	WP1, WP6	The NATWORK trials will be handled in an ethical manner and based on the National and European legislation. The testbed integration access procedure will be planned within NATWORK thoroughly by its GA.

E. Business Risks				
Out of the radar/ emerging competition could hinder innovation and threaten commercialization	High	Med	WP7	Market intelligence activities will ensure continuous monitoring and analysis of the market and competition landscape / EP will ensure the thoroughness and quality of the resulting reports. The market, the exploitation plan will be updated to reduce the risk and new ways of exploitation will be evaluated.
Interoperability problems between the different components of the NATWORK Framework	Med	High	WP6, WP7	Extensive tests will be carried out for all components separately prior to the official testing and their integration to the NATWORK Framework in order to ensure that they were designed and developed according to the project's needs. In this way the proper cooperation among the different components will be ensured. The care given to interoperability further minimizes the risk of such a situation.

3.3. Risk Identification and Analysis

Identifying risks early in the project cycle is crucial. This involves thorough analysis during the planning stages and regular reassessments as the project progresses. Early identification allows for timely implementation of mitigation strategies.

Each identified risk is analyzed to determine its likelihood and potential impact on the project. This dual assessment helps prioritize risks, ensuring that the most significant threats are addressed promptly.

3.4. Categories of Risks

The following categories of risks are identified:

- General Risks: Related to general project management
- Technological Risks: These involve uncertainties related to hardware, software, integration, interfaces, etc.
- User-Related Risks: Related to the adoption of the technologies developed,
- Legal/Organizational Risks: These include compliance issues, disputes, intellectual property rights violations, organizational changes, and ethical considerations.
- Business Risks: These relate to changes in market demand.

3.5. Mitigation Strategies

Developing robust mitigation strategies during the proposal phase is essential. These strategies should be practical, actionable, and specific to the identified risks. Regular monitoring of risk indicators and environmental factors ensures that the project remains on track. This includes setting up early warning systems and conducting periodic risk reviews. Engaging stakeholders throughout the project enhances risk awareness and ensures that all parties are informed and prepared to act if risks materialize. Maintaining close contact with regulatory bodies and staying updated on policy changes helps the project adapt quickly to new regulations. This proactive approach minimizes the risk of non-compliance and associated penalties. Mitigation strategies should be flexible, allowing for adjustments as new risks emerge or existing risks evolve. This adaptive approach ensures that the project remains resilient in the face of uncertainty.

3.6. Ongoing Risk Management

Conducting regular risk workshops with project team members and stakeholders fosters a collaborative environment for risk identification and management. These workshops can highlight potential issues from different perspectives. Maintaining an up-to-date risk register that records all identified risks, their assessments, and mitigation strategies is crucial. Regular updates ensure that the risk register reflects the current risk landscape. Implementing a robust communication plan that ensures timely dissemination of risk-related information to all stakeholders is essential. Clear communication channels help manage expectations and foster a proactive risk management culture. Providing ongoing training and raising awareness about risk management among project team members ensures that everyone understands their roles in identifying and mitigating risks.

4. Data Management Plan

This section provides the Data Management Plan (DMP) of the NATWORK, describing the involved data and how findable, accessible, interoperable, and re-usable (FAIR) data is provided. Then, resources (personnel and infrastructure) are allocated for data management.

Project partners will use this DMP as a compulsory reference for data management within the project, meaning that each project partner will be responsible that the generated and handled data is treated according to the provisions of this DMP.

The project partners have been introduced to the DMP and its use as part of WP1 (Task 1.4) and WP7 activities. WP1 is responsible for addressing any questions and concerns from partners. The DMP can be updated from its creation to the end of project whenever required.

4.1. Data used and generated in NATWORK

The structure of DMP follows the parameters to be clarified regarding the management of the project's generated data. Following the template recommended by the EC [2], the Data Management Plan (DMP) includes the following major components:

- FAIR Data
- Other research outputs
- Allocation of resources
- Data security

It is the responsibility of each consortium partner to ensure the data generated by the partner is treated according to the details laid out in this DMP.

The approach proposed in the current DMP is based on an initial assessment of data handling procedures and anticipated datasets. Each consortium partner has filled-in the questionnaire of data handling presented in Table 2.

Table 2: Questionnaire on data generation, application and management.

Question numbers	Questions
Q1	Will you generate or re-use any existing data and what will you generate/re-use it for? State the reasons if re-use of any existing data has been considered but discarded. What is the expected size of the data that you intend to generate or

	re-use?
Q2	How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)? Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.
Q3	Describe all relevant data quality assurance processes you will apply to ensure the validity and re-usability of the data generated in NATWORK.
Q4	In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g., software, workflows, protocols, models, etc.) or physical (e.g., new materials, antibodies, reagents, samples, etc.). Describe what other research outputs are planned to create in the project and provide sufficient detail on how these research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.
Q5	What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)
Q6	What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?
Q7	Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
Q8	Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?
Q9	Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

The information and data that is expected to be generated by the NATWORK consortium includes the following three types of data:

1. Human readable documents: project documentation, research documentation, publications etc. (typical office document formats, e.g., primarily in the 'Portable Document Format' (PDF)).
2. Source code and software binaries: for implementation of functionalities, support of experimentation and simulation (plain text, software binary formats).
3. Experimental data: raw data from experiments and processed statistical data (raw data).

Regarding the experimental data, during the project lifetime, data will be reused from other sources (e.g., external to the project and/or already available; some examples mentioned in the filled questionnaire in Annex A) and new data will be generated.

In the following sections, the measures followed for providing FAIR data and research outputs are described in detail. Information about the allocation of resources and data security are provided in Annex A, in the description of the plans and measures taken individually by each partner.

4.2. FAIR data principle

The NATWORK consortium follows the principle of FAIR data and so this section describes how the project makes sure its data is findable, accessible, interoperable and reusable. Interested parties might be internal project researchers, or the public in general.

4.2.1. Making NATWORK data Findable

To make NATWORK data findable, standard dataset names with metadata and dataset descriptions will be provided. Datasets generated in NATWORK will be evaluated to determine whether they can be published open access. Where such an evaluation is positive, datasets will be made available online via research data-sharing platforms like Zenodo, which is identified as the primary option for publishing data in NATWORK. In addition to the standard metadata, Zenodo includes the assignment of a DOI, sophisticated versioning, licensing, and access control.

Metadata applied to documents ensures information accuracy and simplifies document search and retrieval. The associated metadata will contain the keywords related to the dataset. Document metadata is information assigned to a document to provide additional context, including characteristics such as what the document is, who created it, and when it was created.

At the end of the project, the coordinator will create a static copy of the web site of the project, including the information on all Open Science activities, storing it on the Internet Archive and on CERTH's servers/cloud used for these purposes. Datasets used for training AI will be made discoverable through Zenodo ElasticSearch and accessible depending on the level of confidentiality associated to the dataset.

We will tag our results using the Metadata format used by Zenodo, which can be exported as MARCXML, Dublin Core, and DataCite Metadata Schema. We will also consider the creation of

an OpenAire Community Gateway for the research associated, together with the rest of funded projects.

For code storage, we will use a hybrid approach. We will store the code currently under development in a temporal internal repository. We will store each release in Zenodo and eventually in a public GitHub repository. Note that Zenodo can only store the code repository's immutable version (a snapshot).

4.2.1.1. Dataset naming conventions

NATWORK will generate multiple data types with different datasets and requirements. However, the datasets of all data types will follow a common naming scheme. The suggested naming scheme is NATWORK_[Type]_[Name]_[Date]_[Partner]_[Counter]_[Version], where:

- [Name] is a short and expressive name for the data;
- [Type] is the type of data, as described in the previous section (i.e., publication, code, design, experimental);
- [Date] is the date when data was produced (format: YYYYMMDD);
- [Partner] is the organization's short name associated with the dataset, as per the partner acronyms at the beginning of this document.
- [Counter] allows to perform multiple experiments/generate multiple datasets per day. It must start at 00.
- [Version] is the numbering of versions of the data;
- _ (underscore) is used as the separator between the fields.

For example, the following name of a potential dataset: NATWORK_experimental_example-dataset_20240617_ELTE_00_v1 would identify the "example-dataset" dataset with an experimental data type. ELTE generated the dataset on 17-06-2024 (and is thus its owner), and the current dataset is the first version.

4.2.1.2. Dataset description and metadata

Each dataset consisting of one or multiple files will be associated with metadata and a dataset description. A dataset may combine files of multiple data types. The description and metadata of the dataset serve the purpose of identification, description, and guidance for use. They will at least contain the following information, fulfilling the requirements of the Dublin Core metadata schemas/standards:

- Identifier(s): an identifier, optionally, additional unique identifiers (e.g., DOI for publicized datasets, a deliverable number of project deliverables etc.).
- Creator/Author: name(s) and affiliation(s).
- Title: title of the dataset.
- Publisher: the entity publishing the dataset.
- Date: the year and month of publication, optionally the day and time of day.
- Type: the dataset type, as defined previously.
- Format: the format(s) of the included data.
- Description: a description of the dataset, including its origin and intended use.

4.2.2. Making NATWORK data Accessible

NATWORK will use Zenodo as the main repository for archiving results. Zenodo (https://zenodo.org) is an open data repository for researchers from all disciplines to share and preserve their research outputs, regardless of size or format. Zenodo makes scientific outputs of all kinds citable, shareable, and discoverable for free in the long term. A Zenodo community has already been created, and all results of the project (including publications) will be linked to this community: https://zenodo.org/communities/natwork/

NATWORK will openly share the work-in-progress source code, open issues, and roadmap on GitHub, where internal and external contributions to any aspect are welcomed. All metadata is available under Creative Common CC0 license, and all open content is accessible through open APIs. Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citable.

The project will rely on complementary open access channels including: (i) The release of pre-publication versions through the website (and possibly arXiv.org); (ii) The release of paper presentations in conferences through the website; (iii) The use of social media to provide links to publications and presentation files. Other "lighter" publications, for example, conference and workshop contributions, will also be provided with Open Access, likely through a "green" approach using, e.g., arXiv.org or other appropriate repositories, according to the publication policy of the conference. Similarly, (public) deliverables produced by the project will also be archived in stable repositories (i.e., not limited to the project's lifetime) to ensure the long-term availability of the material. The deliverables may be archived during an embargo period in coordination with the EC office.

This project is committed to applying Open Science (including open source) policies. As such, NATWORK's default policy for sharing results is to make them as much as possible publicly available and reusable. Due to the participation of large companies in the project, some results

will remain closed for some time (embargo) or undefined. This will be defined on a per-result basis. Currently, for the specific results needed, we aim to determine an embargo time of 6 months after the finalization of the project to allow the protection of all results. All project results will be available in Zenodo (commitment to have a copy of the result in Zenodo two months after generating the result for non-confidential cases). NATWORK will also publish the research process, leading to the result on the project web page in a period shorter than a month after the result is generated. In case, some results (dataset, code, etc.) cannot be made public due to exploitation reasons, metadata will be generated and stored in Zenodo for identification.

NATWORK will also adopt mandatory, well-established open science practices supported by the EC. It will thoroughly follow all the required actions to align with the Open Science practices as defined in the Horizon Europe guidelines. FAIR principles will be applied to our results.

The project will also take advantage of the broad experience of Zenodo to make and maintain project results findable. At the end of the project, the WP7 leader will create a static copy of the project's website, including the information on all Open Science activities, storing it in the Internet Archive and on the project's servers/cloud used for these purposes. Datasets used for training AI will be made discoverable through Zenodo ElasticSearch and accessible depending on the confidentiality level associated with the dataset.

Data will be shared using standard formats. So far, we have not identified the need for any specific data access software. Parts of the code generated by NATWORK will be released by using open-source licenses (e.g., CC, Apache 2.0, etc.).

4.2.3. Making NATWORK data Interoperable

In NATWORK, the consortium is committed to making the generated data interoperable to the maximum degree possible without impacting the execution of work and research within the project and, where possible, with reasonable effort. Standard formats, vocabulary and document structures will be applied to make data exchange possible and simple. Zenodo uses JSON Schema as internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML. For certain terms, we refer to open, external vocabularies, e.g., license (Open Definition), funders (FundRef), and grants (OpenAIRE). We do not anticipate the use of any project-specific ontology or vocabulary.

The data will be properly referenced by scientific publications to allow the replication of the experiments and increasing the impact of the datasets. In addition, the research data will be

made publicly available through the Zenodo open-access repository developed under the European OpenAIRE program.

4.2.4. Making NATWORK data Reusable

NATWORK aims to increase data reuse by publishing everything in standard formats and providing instructions to recreate the experiment or re-process the data. Documentation on how to reuse the project results will be available in the Zenodo community per selected result. When possible, we will provide the scripts (e.g., Python scripts, Jupyter notebooks, etc.) used to process project data or replicate the experiments on both Zenodo and GitHub repositories. Depending on the nature of the result, we will use (as much as possible) open licenses such as EPL-v2.0 or CC-BY. The data published by the NETWORK project will be made available through recommended dataset repositories (Zenodo) to ensure long-term access to the data. For each published dataset, README files and documentation to reproduce and reuse data will be made available and linked with the datasets (stored on Zenodo and GitHub).

The Data & Ethics and GDPR Manager will review on a periodic basis that all results are made available in Zenodo and/or GitHub.

4.3. Other research outputs

Basically, all NATWORK research outputs, including source code, are subject to being included in Zenodo. The working copy of the source code will be stored in a GitHub repository. Once a certain release is ready to be shared with the public, we will link it to Zenodo, creating a specific static copy of it in the GitHub repository.

Finally, we will follow the FAIR principles for all project outputs, specifically including information on the dissemination level intended, the mandatory metadata, and specific metadata, including instructions on how to replicate the results.

5. Ethics Manual

The NATWORK project will consider ethical principles and legal frameworks on Artificial Intelligence (AI) as part of the project life cycle and will implement them according to their specific relevance to the project work. Our research will focus on AI mechanisms applied to security scenarios in various layers from radio to network management. Importantly, no personal data will be utilized in this research. This approach ensures compliance with data protection laws and mitigates any potential ethical concerns related to personal data handling.

5.1. Handling Sensitive Data

If potentially sensitive data is accessed, it will be managed in strict adherence to Horizon Europe's Ethics and Data Protection guidelines. This includes full compliance with the General Data Protection Regulation (GDPR) [3] as well as any applicable national legislation or related EU measures. These practices ensure that all data handling within the project is both legally compliant and ethically sound.

5.2. Ethical Considerations

The project does not foresee any ethical issues arising from the algorithms developed, as they will neither interact with humans nor process human-related data. Additionally, the software produced by the project is not anticipated to cause harm.

5.3. Compliance with Trustworthy AI Guidelines

The NATWORK project is committed to adhering to the European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence. We will ensure compliance with the seven key requirements for Trustworthy AI:

- Human Agency and Oversight: Ensuring that AI systems support human decision-making and uphold human autonomy.
- Technical Robustness and Safety: Developing AI systems that are secure, reliable, and resilient.
- Privacy and Data Governance: Maintaining robust privacy protections and high standards of data management.

- Transparency: Promoting transparency in AI systems and their operations.
- Diversity, Non-discrimination, and Fairness: Ensuring AI systems are inclusive and do not reinforce biases.
- Societal and Environmental Well-being: Creating AI systems that benefit society and the environment.
- Accountability: Establishing mechanisms for accountability and auditability of AI systems.

Our operational approach to achieving these requirements will involve using the Assessment List for Trustworthy AI, developed by the European Commission's High-Level Expert Group on Artificial Intelligence.

5.4. Collaboration with Ethical Committees

Project partners will collaborate with their national or internal ethical committees and/or Data Protection Officers (DPOs) to ensure that all research activities comply with Horizon Europe ethics rules and standards, as well as national guidelines in the partner countries. This collaboration ensures a comprehensive approach to ethics and compliance, fostering a culture of responsibility and ethical integrity throughout the project.

5.5. Continuous Ethics Assessment

The Data & Ethics and GDPR Manager will be responsible for addressing all ethics-related issues. This will involve:

- Continuous Assessment: Regularly evaluating the compliance of research practices with established ethical standards.
- Corrective Actions: Implementing corrective measures whenever necessary to address any deviations or ethical concerns.

5.6. Detailed Procedures for Ethics Management

The Data & Ethics and GDPR Manager will also address all Ethics issues. This will be done by continuously assessing the compliance of the research practices and taking corrective actions.

- Initial Ethics Review: At the project outset, a thorough review of potential ethical issues will be conducted, focusing on data handling, AI implementation, and compliance with relevant guidelines and regulations. See the questionnaire in Annex A.
- Ongoing Monitoring: Throughout the project, continuous monitoring of data handling practices and AI development will be maintained to ensure ongoing compliance with ethical standards.
- Documentation and Reporting: All ethics assessments, reviews, and corrective actions will be thoroughly documented and reported to relevant stakeholders, including ethical committees and funding bodies.

The NATWORK project is dedicated to maintaining the highest ethical standards in all its activities. By adhering to the relevant ethical principles, legal frameworks, and guidelines for trustworthy AI, we aim to conduct our research responsibly and ethically, ensuring that our work contributes positively to society and advances the field of AI in a manner that respects privacy, promotes fairness, and upholds human dignity.

6. Other issues

The project will not use any other procedures for the management of data apart from the ones listed in this document.

7. Conclusions

The current document contains major documentation regarding the assurance of quality procedures throughout the duration of the NATWORK project:

- The Quality Management Plan (Section 2), containing guidelines for ensuring that the project outputs are of high quality.
- The Risk Assessment Plan (Section 3), containing risk identification and mitigation procedures.
- The Data Management Plan (Section 4), containing guidelines for providing FAIR data and project results.
- The Ethics Manual (Section 5), containing guidelines for ensuring the adherence of AI development to ethical principles.

The current deliverable is meant to be used as a reference for the project partners and the relevant management committees of the project, regarding issues of quality of project outcomes and deliverables, mitigation of risks and problems that arise during the project, management of data produced by the project, and ethical considerations in the application of AI for addressing the objectives of the project.

References

[1] AGA, “Annotated Grant Agreement: V1.0 DRAFT– 01.04.2023,” EU Grants:, 2023.

[2] EC, “Data Management Template (HE):V1.0,” EC, 2021.

[3] GDPR, “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data,,” 2016.

Annex A DMP Questionnaire

A.1 Data generated and used by partners

Q1	Will you generate or re-use any existing data and what will you generate/re-use it for? State the reasons if re-use of any existing data has been considered but discarded. What is the expected size of the data that you intend to generate or re-use?
CERTH	CERTH will generally use both newly collected/generated data from our testbeds, but also data generated within the NATWORK project, along with appropriate anonymization modules if required, to enable training of AI models and contribute to the development of various services in the context of the project. As a preliminary step, we will use data and code from open source initiatives, and public datasets, as well. We will also re-use existing data, especially our own historical data and third-party data related to threats and attacks, RAN and UPF traces. The expected data size ranges from 50GB to 500GB.
GRAD	We will use a radio-signal dataset for ML training purposes. Data will come from public sources and will also be generated by Gradiant. It is expected that the dataset can grow up to 1 TB.
TSS	We will generate time series from software control flow tracing. The size of the data in the range between 10MBs to 10GBs. At this point of time, we do not intend to use open-source data but we may reconsider this statement along with the progress made on our AI based software-based detection of DoS and technical exchange with partners involved in this pathway (i.e., MONT) to reach outcome of higher representativity and usability.
CNIT	We will mainly use data and code from open source initiatives, public datasets and publicly available generators. If needed we will deploy new data generation with anonymized traffic traces.
ZHAW	We will use data and code from open-source initiatives and public datasets. We will also generate experimental data from our container-based infrastructure and NATWORK-related testbeds.
UEssex	We will primarily generate experimental data from container-based emulators of end-users/-devices demand and/or benchmarking applications such as iperf and Vegeta. Subject to progress in our AI innovation within the project, the data can be timeseries or event-based with an estimated maximum size of 10GB - open-source data may be used to support experimental data and improve the AI decision support.
ISRD	We will use data from open source initiatives as well as generate our own data in our internal RAN testbed and emulator. The data will be used for training ML models used in solutions developed in the project.

ELTE	We will probably reuse some data and code from open-source initiatives. These include open and anonymized traffic traces to validate our data plane components or training data for validating the low-power AI/ML methods to be developed in the project. During the project we will generate and collect data in the data plane. This data will mostly be used to evaluate the proposed methods under various network conditions (e.g., SLA violations, QoS characteristics, KPIs, routing information, traffic traces, etc.). To support the validation of NATWORK results in a reproducible way, we plan to publish related datasets under an open-source license in public repositories. The generated datasets may vary according to their time granularity: (compressed time-series) 10MBs to 10GBs (traffic traces).
MONT	We will use a combination of open-source datasets and self-generated data, focusing on network traffic for AI-based anomaly detection. Open-source datasets provide diverse traffic patterns essential for training and benchmarking, while self-generated data fill gaps in specific scenarios and recent network threats. The expected data size ranges from 10MB to 10GB, accommodating the needs for training, testing, and validating our AI models to ensure their robustness and accuracy.
IMEC	We expect to generate benchmarking data to use in evaluations. <100MB of data in total.
UZH	UZH intends to leverage UPF traffic traces, RAN traces, and behavioral profiles within the context of various attack scenarios against a 5G network in the NATWORK project. The datasets, which are projected to reach several gigabytes in size, will be systematically analyzed to understand and mitigate discovered security threats. By examining these comprehensive traces and profiles, including existing IoT behavioral profiles, UZH aims to enhance the detection and response mechanisms for potential vulnerabilities in 5G networks.
NEC	Yes, we will use CTI data coming from the Cyber Thread Alliance and other online sources.
NOVA	NOVA is responsible for 'WP7 - Dissemination of Results, Exploitation & Standardization', so only data related to NATWORK's outcomes including technical results and scientific findings as well as news related to NATWORK will be used.
PNET	p-NET's primary role in the project is to support pilots and field trials by offering advanced experimentation facilities and assisting throughout the entire life cycle of experiments and trials. p-NET does not intend to reuse any data. Instead, based on the requirements of the pilots and trials, it will generate datasets to support the development and validation activities. Such datasets may include network traffic data, topology information, infrastructure's energy consumption, etc. The expected size of the data in the range between 10MBs to 10GBs.

A.2 Supporting reproducible research

Q2	How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)? Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.
CERTH	Readme files and documentation will be available on the project repositories to validate data analysis and facilitate data re-use. This includes readme files detailing our methodology, codebooks, data cleaning procedures, analyses, variable definitions, and units of measurement. As a preliminary step, we will re-use data and code from open source initiatives, and public datasets, to enable initial formulation of AI models. We will also re-use existing data, especially our own historical data and third-party data related to threats and attacks, RAN and UPF traces. CERTH will also host the federated repository for boosting AI-powered 6G security, which will enable the collection and the exploitation of the NATWORK autonomous/decentralised data sources in a transparent way. The federated repository will facilitate AI research conducted within NATWORK, but also establish an AI-related federated dataset ready to boost future AI research in B5G/6G networks far beyond the end of this project, towards a European excellence in AI-powered telecommunication infrastructure.
GRAD	Readme files will be available on the project repositories.
TSS	Readme files will be available on the project repositories.
CNIT	Readme files will be available on the project repositories.
ZHAW	Readme files and relevant documentation will be available in the data and code repositories.
UEssex	Documentation will be made available in the project repositories, including (but not limited to): Readme files, Howto files and/or example setup.
ISRD	Readme files will be available on the project repositories.
ELTE	Readme files and documentation will be available on the project web site, data and code repositories.
MONT	We will provide comprehensive documentation to validate data analysis and facilitate data re-use. This includes readme files detailing our methodology, codebooks, data cleaning procedures, analyses, variable definitions, and units of measurement. We plan to re-use existing datasets, such as those from the Canadian Institute for Cybersecurity (https://www.unb.ca/cic/datasets/), for their established reliability and diverse traffic patterns, essential for training and benchmarking our AI models.
IMEC	We will provide documentation, analysis code, and graph generation code under an open source license.

UZH	To ensure robust documentation for validating data analysis and facilitating re-use in NATWORK, UZH will thoroughly compile comprehensive README files, detailed codebooks defining variables and units of measurement, and logs outlining the data cleaning process. Each dataset will be accompanied by thorough documentation of analytical methodologies, including scripts and algorithms used for statistical analyses. We plan to re-use existing datasets, particularly on IoT behavioral profiles, to bolster the reliability and comparability of our findings. Any decision not to use existing data will be clearly justified based on considerations such as data quality, relevance to project objectives, or compatibility issues. This structured approach aims to ensure transparency, reproducibility, and accessibility of our research outputs for future studies and collaborations within the scientific community.
NEC	We will use the standard STIX format for the sharing of the data. The official STIX specification can be found in https://oasis-open.github.io/cti-documentation/stix/intro.html
NOVA	NOVA will not provide documentation regarding the validity of technical data as only NATWORK's outcomes and related news of the NATWORK project will be used.
PNET	Readme files will be available on the project repositories.

Q3	Describe all relevant data quality assurance processes you will apply to ensure the validity and re-usability of the data generated in NATWORK.
CERTH	Data quality assurance processes will be employed across all phases of the project to guarantee that data is accurate, reliable, and accessible for future research and development activities. Readme files and documentation will contain any details on the data generation and/or collection process including the testbed setup, simulation environment, tools, applied methodology and the relevant settings. Detailed validation and verification checks will be conducted also to detect and address any inconsistencies, missing data, or outliers. Additionally, we will employ secure storage, backup, and standardized formats for data sharing, to ensure the validity and re-usability of the data.
GRAD	Readme files will provide information about how to ensure validity and reusability of generated data.
TSS	The provided readme documentation will provide information about data creation and reusability.
CNIT	The provided readme documentation will provide information about data creation and reusability.
ZHAW	Readme files and documentation will contain sufficient information on the data

	generation and/or collection process including the testbed setup/configuration, simulation environment, employed tools, and applied methodology.
UEssex	The readme documentations will provide details on data generation and resuability.
ISRD	Readme files and documentation will contain any details on the data generation and/or collection process including the testbed setup, simulation environment, tools, applied methodology and the relevant settings.
ELTE	Readme files and documentation will contain any details on the data generation and/or collection process including the testbed setup, simulation environment, tools, applied methodology and the relevant settings.
MONT	To ensure the validity and re-usability of the data generated in NATWORK, we will implement comprehensive data quality assurance processes. These include rigorous data collection and preprocessing, detailed documentation (such as readme files and codebooks), validation and verification techniques, and anomaly detection. Additionally, we will maintain secure storage, backup, and standardized formats for data sharing. These measures will ensure that our data is accurate, reliable, and accessible for future research and development activities.
IMEC	Readme files and documentation will contain any details on the data generation and/or collection process including the testbed setup, simulation environment, tools, applied methodology and the relevant settings.
UZH	To ensure the validity and re-usability of the data generated in NATWORK, rigorous data quality assurance processes will be implemented across all phases of the project. This begins with data collection protocols that define clear procedures for obtaining UPF traffic traces, radio traces, and IoT behavioral profiles using standardized methods and reliable instrumentation. Prior to analysis, thorough validation and verification checks will be conducted to detect and address any inconsistencies, missing data, or outliers. Detailed documentation of data cleaning procedures will accompany each dataset, outlining every step taken to preprocess and enhance data quality. Throughout the analysis phase, quality control measures will be applied to monitor the integrity of results, including sensitivity analyses and validation against alternative methodologies. Comprehensive metadata will be provided for each dataset, documenting its origins, variables, definitions, and collection methodologies to facilitate transparent and accessible data re-use.
NEC	We will strictly evaluate the data generated during the NATWORK project using automatic systems. To this end we will establish a process where the obtained information will be compared with that obtained by humans.
NOVA	NOVA will re-share information regarding the status of NATWROK project, including key facts, news related to the project, events, and training and educational sessions. Apart from that, no other data will be created through the communication and dissemination plan. As a result, no validation of these data should be performed.

PNET	Readme files will provide information about how to ensure validity and reusability of generated data.

A.3 Further research outcomes

Q4	**In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g., software, workflows, protocols, models, etc.) or physical (e.g., new materials, antibodies, reagents, samples, etc.). Describe what other research outputs are planned to create in the project and provide sufficient detail on how these research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.**
CERTH	CERTH will make public the results obtained through publications to the scientific community. Datasets and software developed in NATWORK will be stored in a code repository under an open-source license. These results will adhere to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) by ensuring they are documented comprehensively with clear descriptions of their functionalities, versions, and dependencies. In addition, publications will be uploaded to Zenodo (and/or other archive repositories) for accessibility and dissemination.
GRAD	GRAD will make public the results obtained through both publications to the scientific community (uploaded to Zenodo) and patents. Datasets will be publicly available.
TSS	Datasets will be publicly available. Some software will be open source, some other will not be open (e.g., control flow-based time series extraction, Web Assemblies security).
CNIT	Datasets will be publicly available. Some software will be open source, some other will not be open (e.g., DPU development).
ZHAW	ZHAW will make its project results publicly available through both open-access publications and open-source software projects. Relevant suitable datasets will also be publicly available.
UEssex	Datasets will be made publicly available and majority of the research outputs will be open-sourced. Neverthless, UEssex might seek to protect some innovation items if there is a commercialisation/exploitation opportunity.
ISRD	Datasets will be publicly available. Research results will be available via open-access publications. Software will be available through commercial licensing.
ELTE	Datasets will be assigned to different components and methods developed in NATWORK. For each dataset, we will provide detailed descriptions as README.MD files that can be used to understand and analyze the data and to

	reproduce the experiment in which it was created. We will also provide script or Jupyter notebooks for data preprocessing and analysis, depending on the nature of results. Software developed in NATWORK will be stored in a code repository under an open-source license. Publications will be uploaded to Zenodo.
MONT	Datasets will be publicly available with comprehensive descriptions in README.MD files, enabling users to understand and analyze the data and reproduce the experiments in which it was generated. All software created within NATWORK will be housed in a code repository under an open-source license. Furthermore, publications will be uploaded to Zenodo for accessibility and dissemination.
IMEC	Software developed in NATWORK will be stored in a code repository under an open-source license. Snapshots of the software used in publications will be uploaded to Zenodo.
UZH	In addition to managing data in NATWORK, UZH anticipates generating various other research outputs that will contribute to the advancement of knowledge and support reusability. These outputs encompass the digital form. We plan to develop software tools for data analysis, workflows for processing UPF traffic and radio traces, and protocols for conducting IoT behavioral profiling. These digital outputs will adhere to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) by ensuring they are documented comprehensively with clear descriptions of their functionalities, versions, and dependencies. They will be made accessible through repositories and platforms commonly used in our research community, facilitating easy discovery and reuse by other researchers.
NEC	NEC will make public the result obtained through both publications to the scientific community and patents.
NOVA	NOVA is responsible for 'WP7 - Dissemination of Results, Exploitation & Standardization', so no data related to technical parts of NATWORK will be generated and / or reused.
PNET	PNET will make public to the scientific community any new results through publications and patents. If new software is developed, open-source license will also be considered.

A.4 Cost of data management and data sharing

Q5	**What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)**

CERTH	Diverse costs are related to storage, archiving, re-use, security, and management of data in order to align with FAIR principles. Indirect costs include expenses for dedicated servers capable of securely storing large volumes of data, software, and models at CERTH infrastructure. These costs also involve metadata creation and maintenance. Moreover, direct costs such as personnel effort for data curation, maintenance of repositories, and compliance with data protection regulations are required.
GRAD	Indirect costs will cover this for data stored at GRAD infrastructure.
TSS	Indirect costs will cover this for data stored at TSS infrastructure.
CNIT	Indirect costs will cover this for data stored at CNIT infrastructure.
ZHAW	Indirect costs will cover this for data stored at ZHAW infrastructure.
UEssex	Indirect costs will cover this for data stored at UEssex infrastructure.
ISRD	Indirect costs will cover this for data stored at ISRD infrastructure.
ELTE	Indirect costs will cover this for data stored at ELTE infrastructure.
MONT	Indirect costs will cover this for data stored at MONT infrastructure.
IMEC	Indirect costs will cover this for data stored at IMEC infrastructure.
UZH	Ensuring that our project's data and other research outputs adhere to FAIR principles involves several cost considerations related to storage, archiving, re-use, security, and management. Indirect costs include expenses for cloud storage services or dedicated servers capable of securely storing large volumes of data, software, and models. These costs also encompass metadata creation and maintenance, ensuring all outputs are adequately documented and accessible. Additionally, direct costs such as personnel time for data curation, maintenance of repositories, and compliance with data protection regulations are essential. In NATWORK, UZH will provide its infrastructure to store data free of charge. This includes utilizing UZH's servers and storage facilities (e.g., Nextcloud[1]) and ensuring NATWORK data is safely archived and accessible throughout the project. However, costs associated with documenting data, including the creation of metadata, will be covered as part of the project expenses. This allocation ensures that our research outputs are comprehensively documented by FAIR principles, promoting their discoverability, accessibility, and reusability by the broader research community. [1] https://owncloud.csg.uzh.ch/
NEC	Indirect costs will cover this for data stored at NEC infrastructure.
NOVA	NOVA's infrastructure will cover any direct and indirect costs regarding data storage and archiving.
PNET	Indirect costs will cover this for data stored at PNET infrastructure.

A.5 Data security

Q6	What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?
CERTH	CERTH will host the federated repository for boosting AI-powered 6G security, which will enable the collection and the exploitation of the NATWORK autonomous/decentralised data sources in a transparent way, utilizing access control mechanisms to restrict data access to authorized personnel only. CERTH does not expect to generate or use sensitive data. In case that we generate sensitive data, appropriate anonymization techniques and encryption protocols will be employed to these data, ensuring that information remains protected against unauthorized access or interception. For non-sensitive datasets, we will use Zenodo or GitHub as certified repositories for ensuring long-term availability alongside regular backups and recovery protocols at CERTH infrastructure in order to ensure data recovery capabilities and mitigate risks of data loss. Security measures, such as firewalls and intrusion detection systems will be deployed also to strengthen the security of the data storage.
GRAD	No sensitive data will be part of the radio-signal database. GRAD's backups will ensure data recovery in case of loss/corruption.
TSS	No sensitive data is expected from our control flow time serie extraction. These data are too specific to the sample software we will monitor and therefore cannot be exploited maliciously.
CNIT	Sensitive data will be privately stored in CNIT repositories. We do not plan to use them.
ZHAW	Any sensitive data will be privately stored in ZHAW servers and digital infrastructure. ZHAW already has rules and procedures as a Swiss university in place to securely manage such data and implements secure backup, archiving, and recovery workflows.
UEssex	UEssex does not plan to develop or interact with real sensitive data, instead might generate fake sensitive data (i.e. synthetic/artificial data under sensitive features) that does not lead to real users or end-devices. Where container identity need to be protected, anonymization and/or filtering methods will be applied to the datasets before sharing them outside UEssex infrastructure.
ISRD	No sensitive data will be produced or used in this project. Backup procedures will be in place to prevent data loss.
ELTE	Sensitive data will always be stored in private repository. For non-sensitive datasets, we will use Zenodo or GitHub as certified repositories for ensuring long-term availability. Note that we do not plan to use sensitive data in this project.
MONT	No sensitive data is expected. Regular backups will ensure data recovery in case

	of loss or corruption. Additionally, data transfer will be secured using encrypted channels to prevent unauthorized access. Access controls and authentication mechanisms will be in place to restrict data access to authorized personnel only, ensuring the integrity and confidentiality of our data throughout its lifecycle.
IMEC	We do not expect to generate or use sensitive data. For non-sensitive datasets, we will use Zenodo or GitHub as certified repositories for ensuring long-term availability.
UZH	In NATWORK, ensuring robust data security is a top priority to safeguard the confidentiality and integrity of our research outputs. Data will be securely stored on dedicated servers provided by UZH, utilizing access control mechanisms such as Access Control Lists (ACLs) to restrict data access to authorized personnel only. Encryption protocols will be applied to sensitive data both at rest and during transfer, ensuring that information remains protected against unauthorized access or interception. Comprehensive data backup and recovery procedures will be in place to mitigate risks of data loss. Importantly, UZH confirms that no sensitive data collection is planned for this project, aligning our practices with careful data protection regulations and ethical guidelines. These measures collectively reinforce our commitment to maintaining the highest standards of data security throughout the project duration, fostering trust and compliance with legal and ethical standards in data handling.
NEC	No sensitive data is expected to be used during the NATWORK project.
NOVA	NOVA will not use any sensitive data for dissemination and communication purposes. Any shared information will be placed to the common repository.
PNET	No sensitive data is expected to be used during the NATWORK project. Considering data security (secure storage and recovery), PNET cloud infrastructure also has in place processes for them (backups, access control etc.).

A.6 Data ethics

Q7	**Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).**
CERTH	Currently, there are no ethical/legal issues as we are not using any sensitive data. In the case that we have access to such data in the future e.g., pertaining to users' mobility etc., data will be treated according to Horizon Europe Ethics and Data Protection guideline by considering GDPR as well as national legislation or related EU measures.

GRAD	No relevant ethical/legal issues are envisioned.
TSS	No relevant ethical issues are envisioned.
CNIT	No relevant ethical issues are envisioned.
ZHAW	No relevant ethical/legal issues are envisioned.
UEssex	No relevant ethical/legal issues are envisioned.
ISRD	No relevant ethical/legal issues are envisioned.
ELTE	Currently, there are no ethical/legal issues as we are not using any sensitive data. In the case that we have access to such data in the future e.g., pertaining to users' mobility etc,, data will be treated according to Horizon Europe Ethics and Data Protection guideline by considering GDPR as well as national legislation or related EU measures.
MONT	There are no evident ethics or legal issues impacting data sharing. All data collection and sharing processes comply with relevant legal and ethical standards. Any potential concerns will be addressed through continuous adherence to established guidelines and thorough ethics reviews.
IMEC	Currently, there are no ethical/legal issues as we are not using any sensitive data. In the case that we have access to such data in the future e.g., pertaining to users' mobility etc,, data will be treated according to Horizon Europe Ethics and Data Protection guideline by considering GDPR as well as national legislation or related EU measures.
UZH	In NATWORK, potential ethics and legal considerations regarding data sharing primarily revolve around ensuring compliance with data protection regulations and ethical guidelines. Specifically, UZH confirms that no sensitive information will be collected, which significantly mitigates risks associated with privacy and confidentiality. This approach aligns with the ethical review process outlined in our project's Description of the Action (DoA), ensuring adherence to established guidelines and principles for responsible data management.
NEC	There are no evident ethics or legal issues impacting data sharing. All data collection and sharing processes comply with relevant legal and ethical standards. Any potential concerns will be addressed through continuous adherence to established guidelines and thorough ethics reviews.
NOVA	Currently, there are no ethical/legal issues as we are not using any sensitive data. In the case that we have access to such data in the future e.g., pertaining to users' mobility etc,, data will be treated according to Horizon Europe Ethics and Data Protection guideline by considering GDPR as well as national legislation or related EU measures.
PNET	No relevant ethical issues are envisioned.

Q8	**Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?**
CERTH	CERTH will not collect personal data, thus there is no need for informed consent

	regarding data sharing and long-term preservation through questionnaires.
GRAD	There is no plan to use personal data.
TSS	We do not plan to deal with personal data.
CNIT	We do not plan to deal with personal data.
ZHAW	We do not plan to deal with personal data.
UEssex	We do not plan to deal with personal data.
ISRD	We do not plan to deal with personal data.
ELTE	We do not plan to deal with personal data.
MONT	We do not plan to deal with personal data.
IMEC	We do not plan to deal with personal data.
UZH	In NATWORK, UZH will not collect personal data, thereby obviating the need for informed consent regarding data sharing and long-term preservation through questionnaires. By strictly adhering to this principle, we ensure that privacy concerns are mitigated from the outset.
NEC	We do not plan to deal with personal data.
NOVA	NOVA will not use any personal data throughout the timeframe of the NATWORK project. No personal data will be disseminated, if required only aggregated and anonymised data will be used.
PNET	We do not plan to deal with personal data.

A.7 Other procedures for data management

Q9	Do you, or will you, make use of other national/funder/sectorial/ departmental procedures for data management? If yes, which ones (please list and briefly describe them)?
CERTH	No, we do not make use of other national, funder, sectorial, or departmental procedures for data management.
GRAD	We will not make use of other funding for data management.
TSS	No, we do not plan to make use of other funding for data management.
CNIT	No, we do not plan to make use of other funding for data management.
ZHAW	No, we do not plan to make use of other funding for data management.
UEssex	No, we do not plan to make use of other funding for data management.
ISRD	No, we do not make use of other national, funder, sectorial, or departmental procedures for data management.
ELTE	The project topics where AI is going to be applied will adhere to current regulation in the EU and non-EU countries involved in the project, including the guidelines, standards and plans published by the EU Commission, the Department of Commerce of the US and the National Institute for Standards and Technology. The AI algorithms will only tackle management of resources,

	without requiring human data and not interacting with any other system.
MONT	No, we do not make use of other national, funder, sectorial, or departmental procedures for data management.
IMEC	No, we do not make use of other national, funder, sectorial, or departmental procedures for data management.
UZH	In our project, we will adhere to several key data management procedures to ensure robust and compliant practices within Switzerland. Firstly, we will follow the Swiss National Science Foundation (SNSF) guidelines, which emphasize open data and FAIR principles[1]. These guidelines will shape our data management plan, ensuring proper data sharing, preservation, and accessibility. Secondly, we will comply with the UZH policies, which mandate secure storage, data encryption, and strict access controls to safeguard data integrity and confidentiality[2]. Lastly, our practices will align with the Swiss Federal Act on Data Protection (FADP)[3], which governs the processing of personal data in Switzerland. Although we do not collect personal data, adhering to the FADP principles will further ensure that our data management processes are secure and ethical, maintaining high standards of data protection and compliance. [1] https://www.snf.ch/en/FAiWVH4WvpKvohw9/topic/research-policies [2] https://www.ub.uzh.ch/en/wissenschaftlich-arbeiten/mit-daten-arbeiten/FAIR-und-Open-Data.html [3] https://www.fedlex.admin.ch/eli/cc/2022/491/en
NEC	No, we do not make use of other national, funder, sectorial, or departmental procedures for data management.
NOVA	NOVA will consult the European Commission's Dissemination and Exploitation guidelines and related services available.
PNET	No, we do not plan to make use of other funding for data management.