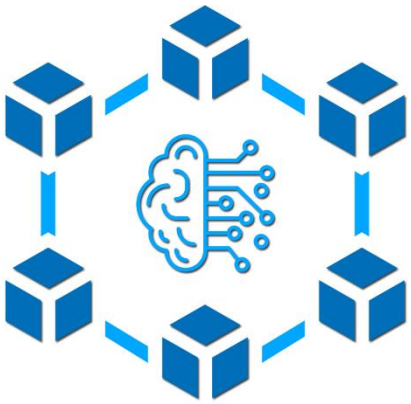

NANCY

An Artificial Intelligent Aided Unified Network for Secure Beyond 5G Long Term Evolution [GA: 101096456]

Deliverable 1.5

Data Management Plan

Programme: HORIZON-JU-SNS-2022-STREAM-A-01-06

Start Date: 01 January 2023

Duration: 36 Months

Co-funded by the European Union

6GSNS

NANCY project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096456.

Document Control Page

Deliverable Name	Data Management Plan
Deliverable Number	D1.5
Work Package	WP1
Associated Task	T1.4
Dissemination Level	Public
Due Date	M06 – 30 June 2023
Completion Date	30 June 2023
Submission Date	01 July 2023
Deliverable Lead Partner	SID
Deliverable Author(s)	Konstantinos Kyranou (SID), Zisis Batzos (SID), Ioannis Chochliouros (OTE), Maria Belesioti (OTE), Christina Dolianidi (DRAXIS), Maria Tzana (SID)
Version	**1.0**

Document History

Version	Date	Change History	Author(s)	Organisation
0.1	08 May 2023	Table of Contents	Konstantinos Kyranou, Zisis Batzos	SID
0.2	30 May 2023	Contributions to Sections 2 and 6	Konstantinos Kyranou, Zisis Batzos, Maria Tzana	SID
0.3	09 June 2023	Contributions to Section 5	Ioannis Chochliouros Maria Belesioti	OTE
0.5	17 June 2023	Contributions to Sections 3 and 4	Christina Doliannidi	DRAXIS
0.8	27 June 2023	Internal Review	Ioannis Choclioupos, Christina Doliannidi	OTE/DRAXIS
1.0	29 June 2023	Revisions – Final Version	Konstantinos Kyranou, Zisis Batzos	SID

Internal Review History

Name	Organisation	Date
Ioannis Chochliouros	OTE	27 June 2023
Christina Dolianidi	DRAXIS	28 June 2023

Quality Manager Revision

Name	Organisation	Date
Anna Triantafyllou	UOWM	30 June 2023
Dimitrios Pliatsios	UOWM	30 June 2023

Table of Contents

List of Figures

List of Tables

List of Acronyms

Explanation	Acronym
5G	The Fifth Generation of Mobile Communications
ABS	Access and Benefit Sharing
ADR	Alternative Dispute Resolution
AI	Artificial Intelligence
ASCII	American Standard Code for Information Interchange
ADFA	Australian Defence Force Academy
BSD	Berkeley Source Distribution
BW	Bandwidth
CA	Consortium Agreement
CERN	*Conseil Européen pour la Recherche Nucléaire*
CSV	Comma-separated values
DMP	Data Management Plan
DB	Database
DBF	database file
DL	Deep Learning
DoS	Denial of Service
DoA	Description of Action
DOI	Digital Object Identifier
DDoS	Distributed Denial-of-Service
EC	European Commission
EU	European Union
FAIR	findable, accessible, interoperable, and reusable
GDPR	General Data Protection Regulation
GPL	General Public License
GA	Grant Agreement
HE	Horizon Europe
HTML	Hypertext Markup Language
HTTP	HyperText Transfer Protocol
IIoT	Industrial Internet of Things
IP	Intellectual Property
IPR	Intellectual Property Rights
ICMP	Internet Control Message Protocol
IoT	Internet of Things
JSON	JavaScript Object Notation
KPI	Key performance indicator
KB/ MB/ GB	Kilo/mega/giga bytes
LGPL	Lesser General Public License
ML	Machine Learning
MIT	Massachusetts Institute of Technology
MCARD-HEU	Model Consortium Agreement for Research, Development and Innovation Actions under Horizon Europe
MPL	Mozilla Public License
MEC	Multi-access edge computing
NDA	Non-Disclosure Agreement
NFV	Network Functions Virtualisation
ODS	OpenDocument Spreadsheet
ODT	Open Document Text

OSI	Open Source Initiative
OSS	Open Source Software
PDF	Portable Document Format
Recon	Reconnaissance
RAN	Remote Access Network
R&D	Research and Development
R&I	Research and Innovation
RDM	Research Data Management
SEIT	School Of Engineering and Information Technology
SME	Small-Medium Enterprise
SW	Software
SQL	Structured Query Language
SYN	Synchronise
TAB	Tab-delimited file
TCP	Transport Control Protocol
TN	Test Network
UDP	User Datagram Protocol
UE	User Equipment
UNSW	University of New South Wales
WP	Work Package
WIPO	World Intellectual Property Organization
XLSX	Excel Microsoft Office Open XML Format Spreadsheet file
XML	eXtensible Mark-up Language

Executive summary

NANCY project's Data Management Plan's initial iteration is presented in this document. The Data Management Plan (DMP) strategy outlines the data management life cycle for the information which the project will gather, manage and/or (re)produce.

According to Horizon Europe's open science requirements[1], proper Research Data Management (RDM) is mandatory for any Horizon Europe (HE) project which aims at generating or reusing research data. In this guide, the process of making research data findable, accessible, interoperable, and reusable (FAIR) is elaborated, so that it would be properly handled. The fundamental channel for scientific discovery and innovation, as well as later data and knowledge integration and reuse, is strongly intertwined with good research data administration. Based on the aforementioned recommendations, a template was developed and provided to all involved parties. Throughout the project, this deliverable will be revised if there are any substantial changes, such as the availability of new datasets and information, consortium regulations, changes in the consortium's constitution, or any other external causes. In the forthcoming periodic reports of the project, any updated or enhancements on the DMP will be also presented. The datasets gathered throughout the project's initial six-month period, are displayed in Annex B.

The very nature of collaborative research and innovation (R&I) initiatives requires that several partners with various perspectives and interests gather around a common table. In light of this, effective knowledge management and protection are necessary to be a key component of the NANCY project's overall administration. The management of Intellectual Property (IP) Assets in accordance with the European Union (EU) and international rules, also falls under the purview of Task 1.4, hence this deliverable includes additional information on that topic.

[1] For more details see: European Commission: *The EU's open science policy*, available at: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

1. Introduction

This section briefly provides high-level information related to the various objectives the present deliverable aims to realise, as well as introduces the liaisons with other WPs and Tasks within the scope of the entire project.

1.1. Scope of the Deliverable

NANCY's Data Management Plan (DMP) is described and presented in this deliverable. The "Template for Horizon 2020 DMP" version 1.0 [1], which was released by the European Commission on February 15, 2018, establishes the rules and regulations that the NANCY DMP adheres to. This procedure was used to develop, characterise, and maintain the NANCY datasets. The NANCY collaborative system (i.e., NANCY Confluence), which is used to support the NANCY DMP, was introduced in the respective Deliverable 1.1 - Project and Risk Management Handbook.

The administration of developed and gathered expertise is the subject of detailed Task 1.4. The dissemination and utilisation of knowledge elements are controlled by explicit policies. The development and implementation of NANCY's DMP in accordance with the General Data Protection Regulation (GDPR[2]) and the European Commission (EC) Guidelines on Data Management in Horizon 2020[3] takes place in the framework of this assignment. The DMP defines the roles in charge of managing the content and metadata of data components, as well as the various procedures, rules, and obligations for managing a legal entity's/organization's full dataset in accordance with legal and/or policy requirements set for such purposes.

A thorough questionnaire was structured according to the instructions in [1] and disseminated to all NANCY partners/beneficiaries who are compiling data, asking them to submit their responses. Partners provided responses based on what they knew at the time this deliverable was generated. The DMP will be revised in due course with the project's first periodic report (scheduled for M18) and second periodic report (scheduled for M36).

This is the initial version of the DMP, which was created at the end of June 2023, the sixth month of the project.

1.2. Connection with other Tasks & Deliverables

WP1 – "Project, Innovation & Data Management" assists with project management, cooperation, and finance management responsibilities. Data processing and Intellectual Property Rights (IPR) management are two of WP1's main activities. Task 1.4 – "GDPR-compliant Data Management" receives input from the Grant Agreement (GA) and produces D1.5 – "Data Management Plan". NANCY's DMP documents and monitors the datasets as well as other intellectual properties throughout the project's lifecycle, providing the appropriate landscape to develop a successful exploitation strategy that will be beneficial for each consortium's participant. There is a connection between the following project tasks:

- Project coordination and management
- Use Case creation and implementation

[2] For more information regarding the Regulation (EU) 2016/679 (General Data Protection Regulation) visit: https://gdpr-info.eu/

[3] For more information regarding the Open Access and Data Management visit: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

- Development along with integration
- Dissemination and exploitation

1.3. Structure of the Document

A concise and understandable comprehension of IPR and dataset preservation methods are provided by the meticulous layout and organisation of the current document. The document is divided into several different sections, each of which is devoted to a particular facet of the subject at hand:

- **Section 2 – Data Management Methodology:** In this section the various aspects of the dataset template are clarified, elaborations are provided regarding the FAIR guidelines, and how the NANCY project complies with each of these concepts is explained. Moreover, the allocated responsibilities are showcased from the DMP perspective and provide information for the security mechanisms.
- **Section 3 – Data Management Plan monitoring & update:** This section gives the provisions regarding the potential updates and enhancements through the project's life cycle.
- **Section 4 – NANCY Datasets Overview:** In this section, a high-level observation of the datasets provided by the partners is displayed.
- **Section 5 – NANCY IPR management:** This part of the document highlights the critical rules that regulate the various IPRs' access rights, potentially affecting NANCY's progress and evolution.
- **Section 7 – Conclusion:** This section concludes the deliverable.
- **Annex A:** Presents the NANCY Dataset Template.
- **Annex B:** Presents the NANCY Datasets compiled from the partners.

2. Data Management Methodology

The NANCY consortium created a template based on the "Template for HORIZON 2020 DATA MANAGEMENT PLAN (DMP)" [1] for clarifying, characterising, and maintaining the numerous NANCY datasets. This template, presented in Annex A, was developed by utilising a seven-features methodology. These seven characteristics include:

i. Dataset Summary
ii. Task and Deliverable Relation
iii. Ethical Aspects
iv. Partners and Services Responsibilities
v. FAIR principles
vi. Resource allocation
vii. Data Security

These characteristics constitute the foundation of NANCY DMP. The NANCY data management approach is shown in Figure 1, while the subsequent sections provide an in-depth look at each of the aforementioned traits.

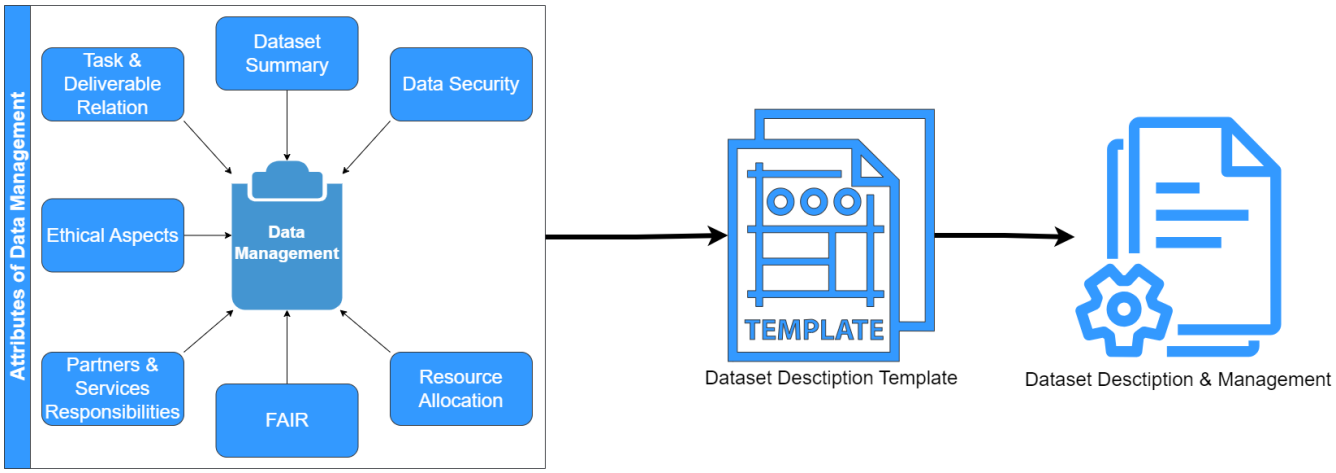

Figure 1: Data Management Methodology of NANCY

2.1. Data Summary

NANCY consortium strives towards prioritising the utilisation of open data when it is feasible while maintaining privacy and personal information protection. It should be noted that information isn't going to be made publicly available if it could compromise an individual's privacy or confidentiality or if it poses a security risk.

The following details are included in the template's relevant dataset summary section:

- Dataset Description: Provide a brief description of the dataset
- Dataset Purpose: Provide specifications regarding the purpose of the dataset
- Dataset Type/Format: Provide the format in which the dataset is provided (CSV file, raw TXT, MongoDB or MySQL, etc.)
- Re-use of existing data: State if the dataset has previously used data
- Dataset Origin: Specify the origination of the dataset
- Dataset Collection: Provide specifications on how the dataset has been collected

- Expected Size: Provide the estimated magnitude of the dataset

2.1.1.Objective of data collection / generation and relevance to the NANCY project

This comprises a variety of data that needs to be acquired to be able to achieve the action's objective. Some of the NANCY-related tasks and objectives that call for data gathering include the following:

- For the purpose of improving the design and construction of the NANCY components, it is crucial to understand the expectations, difficulties, and requirements of end users and involved stakeholders. Instead of depending on hypotheses, the objective is to build solutions that are tailored to the actual demands of the end users.
- In order to develop the final version of the components and tools, it is crucial to track and assess how the NANCY components are being used in the project's use cases, while taking into account the views of both NANCY's stakeholders and NANCY's technical partners.
- To be able to gauge their impact, adjust the NANCY agenda accordingly, and satisfy the Commission's reporting obligations, it is critical to track and assess the project's dissemination and communication results.

On the basis of the aforementioned information, the following data is anticipated to be gathered:

i. Surveys, questionnaires, assessments, and interviews are examples connected to research, innovation, and communication.
ii. Technological information, such as network traffic data, infrastructure data, etc., is expected to be collected in the project's development, implementation, and demonstration phases
iii. Key Performance Indicators (KPIs) regarding the project's efficacy and the communication and dissemination activities are essential in order to assess progress.

The DMP documentation will be revised and enhanced to reflect the new data summaries, forms of consent, compliance, and organisational authorization, as required when private information gathering or analysis takes place. The Data Protection Principles shall be followed while collecting, handling, and analysing personal data. All essential steps will be taken by the management team of the project and by each beneficiary to guarantee adherence to national [2] and European laws, as well as any professional rules of conduct pertaining to the safeguarding of personal data. This will specifically cover the General Data Protection Regulation (GDPR, 2016/679) [3], Directive 95/46/EC [4], and relevant national standards, guaranteeing compliance with laws and regulations.

2.1.2.Types and formats of collected/generated data

As part of the NANCY project, a substantial amount of information will be managed. Each data item has an indication for its characteristics, format, source, estimated magnitude, storage location, function, and potential re-use. The "Type of data" specifies the type of information to which a certain piece of data belongs. The following are only a few examples of the anticipated data types:

- Pre-existing data sets
- Contribution received directly from partners or external stakeholders
- Data acquired during usage scenarios
- Information gathered during workshops or stakeholder collaboration sessions
- Analytics data gathered from social platforms, such as LinkedIn profile and the website
- Information gleaned from public database repositories

The "Format of data" identifies the category or style of support that a particular set of data belongs to. The following are only a few examples of the anticipated data formats:

- JavaScript Object Notation (JSON)
- eXtensible Mark-up Language (XML)
- Comma-separated values (CSV)
- XML-based open file format and proprietary Microsoft Excel format (XLSX/XLS)
- Text-formatted data
- Database file
- Tab-delimited file (TAB)
- Microsoft Word format (DOC/DOCX)
- OpenDocument Spreadsheet (ODS)
- Open Document Text (ODT)
- Hypertext Markup Language (HTML)

2.1.3. Data sources and leveraging of pre-existing data

The source of the dataset should describe the methods used to gather the data as well as the tools that were used. The primary sources of data comprise (but are not restricted to):

- Polls
- Surveys
- Questionnaires
- Requirements / Specifications
- Demonstrations and Tests
- Digital or physical format of use case records

The capacity to reuse data determines the extent to which the information will be regarded as private and, if so, the authorisation level of stakeholders who will be allowed to gain access to it.

2.1.4. Data utility & volume

Each data criterion regarding the magnitude of the datasets is defined in accordance with the list below:

- Expected Size: Indicates the anticipated volume of the data in KB, MB, or GB.
- Storage: This applies to the storage of information repositories, such as Zenodo[4] (for open access to information) or the internal storage of NANCY, where the data will be kept

The term "Data utility" refers to the purpose and aim of data processing and specifies who may utilise the information for purposes other than those of the NANCY project.

2.1.5. Data reference and naming identifiers

Each dataset is uniquely identified for the project purposes, following a specific syntax as shown in Figure 2:

Filename Syntax: NANCY_WP1_Dataset_Template_Organization_name.docx
Example: NANCY_WP1_Dataset_Template_SID.docx

Figure 2: Dataset Naming Convention

A single document for each one of the provided datasets is uploaded to the internal repository of NANCY. Inside the NANCY dataset template document, there is an allocated sub-section at the “FAIR

[4] For further information also see: https://zenodo.org/

Data - Making data findable, including provisions" section, where partners can additionally provide any other naming conventions for the dataset.

2.2. FAIR Data

FAIR stands for Findable, Accessible, Interoperable, and Reusable. It is a set of guiding principles that do not require data producers to adopt specific standards or technologies. Instead, it encourages researchers to think about the bigger picture, where their data fits into the larger context of their research field, and how it can be optimised for impact and longevity.

Using common standards makes it easier for others to find and access your data for their research through reuse. This is all possible through interoperability based on shared metadata and ontologies and an open license in a public repository. The "FAIR Guiding Principles for scientific data management and Stewardship" intend to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reusability (FAIR) of digital assets. These principles refer to data (or any digital object), Metadata (information about that digital object), and infrastructure.

2.2.1. Making data findable, including provisions for metadata

This is the first aspect of FAIR and it refers to making datasets clearly recognisable and identifiable. Each dataset must adhere to the following requirements in order to accomplish this:

- To decide if and in what manner the data may be identified, metadata and special identifiers like the Digital Object Identifier (DOI) will be employed.
- A dataset's name scheme will be determined by the "Name Convention".
- Versioning will be employed to indicate whether or not it is supported.
- In addition to the dataset's filename, history, and other metadata will be utilised.

Zenodo, which is a combined open-source archive of scholarly writings and data managed by CERN[5] and OpenAIRE[6], is used by NANCY. The following format will be used by the NANCY partners when giving access to the publication and identifying the bibliographic metadata [5]:

- Funding Grant: Horizon Europe
- NANCY, Grant Agreement No 101096456
- Publication date, persistent identifier

The most important step in data (re)use is to find it. Both data and Metadata should be easily findable by humans and computers. Machine-readable Metadata is essential for automatically detecting datasets and services, comprising a necessary part of the FAIRification process. There are four principles governing the findability of the (meta)data. (Meta) data receive a globally unique and permanent identifier, and the data is described using extended Metadata. Moreover, Metadata clearly and conspicuously contains the identifier of the data they describe and is stored or indexed in a searchable resource.

2.2.2. Making data openly accessible

Data accessibility, which is related to data availability and access options, is the second tenet of FAIR. For this purpose, fields like the ones that follow are used:

[5] *CERN (Conseil Européen pour la Recherche Nucléaire) is the* European Organization for Nuclear Research and it is one of the world's largest and most respected centres for scientific research. For further information also see: https://www.home.cern/

[6] For further information also see: https://www.openaire.eu/

- "Dataset Openly Accessible" displays the dataset's accessibility option
- If the dataset cannot be made publicly available, further information is also provided.
- Using standardised communication protocols that are open, free, and universally implementable, data may be retrieved by their unique identification, and it allows for a suitable authentication and authorisation method.
- The programme and/or mechanism must be mentioned in order to obtain access to the dataset.
- The repository that belongs to the owner should have the capacity to offer the necessary storage, searching, and accessing capabilities.

Humans and machines can access data via an open, free, and universal standardised communication protocol that allows authentication and authorisation procedures when required. The data must not be opened. Data may be sensitive for confidentiality, national security, or commercial interests. If they cannot be opened, the conditions of access and reuse must be clear and transparent.

2.2.3.Making data interoperable

Interoperability, the third FAIR tenet, describes whether or not the dataset is interoperable as well as the standards or practices that promote interoperability. The classification for each and every one of these categories is the following:

- The interoperability determines whether or not the dataset is interoperable, taking into account the associated elaboration.
- The adoption of several types of techniques will facilitate interoperability in terms of the NANCY project.

The data and metadata norms and techniques used by NANCY are intended to make it easier for project data to be interoperable. NANCY complies with general format standards and is compatible with openly downloadable (open-source) software programmes to make it easier to reuse data from diverse sources.

The data and associated metadata use "a formal, accessible, shared, and broadly applicable language for knowledge representation". This includes using community-accepted languages, formats, and vocabularies in the data and Metadata to follow FAIR principles. Metadata includes qualified references to other (meta) data, Metadata, and information and describes them by using identifiers.

2.2.4.Making data openly reusable

Data re-use, the fourth driving concept of FAIR, aims to render produced data as accessible as feasible. Therefore, specific data licences, sharing agreements, and due dates are specified in writing. There are several distinct categories used, as follows:

- If a particular licence (such as Creative Commons) is followed, it dictates how the data is acknowledged
- The conditions for the dissemination of open data are laid forth in a box titled "Data Sharing Terms." The sharing terms are implemented when there is no explicit licence in place
- Whether or not information will be kept publicly accessible after the NANCY project is complete, is indicated by the "Dataset Sharing after the Project's End" field
- If the dataset will be updated once the NANCY project is over, may be seen in the section under "Data preservation and updating after the project"

- Users can determine the length of time during which the dataset should be accessible via the "re-use timescale"

Data from the pilots will be connected to research throughout the project's life cycle, with the potential to protect business or industrial interests. Such use case information is not commonly accessible for assessment and reuse, due to intellectual property protection requirements. The data is going to be published in gold open access or green open access[7], depending on the Consortium members' permission. Information under open access on other platforms will be subject to repository regulations, such as the Zenodo storage regulations [6]. When releasing information, the data provider or creator is responsible for its integrity.

When feasible and after considering any potential personal information conflicts, intellectual property rights, security concerns, and any other pertinent legal and ethical criteria, open access licenced data must be utilised. The project will make every effort to make it possible for other parties to reuse data based on Open Access standards, to the degree that IPR or other rights demand such constraints. Information regarding data licencing will be available from OpenAIRE Research Data Incense [7]. All study-related data will be preserved for two years following the conclusion of the research project, in line with the requirements mentioned, until it may be utilised again. A data-sharing agreement will be created in accordance with the terms, which will outline the manner in which data will be anonymized or summarised, the participant's consent for data sharing, copyright permissions, and consensus on any prohibition periods if data is to be restricted or shared with third parties outside of the consortium.

The ultimate goal of the FAIR methodology is to ensure an optimised reuse of data. To achieve this goal, the associated metadata provide comprehensive and accurate information, as well as render the data to be accompanied by an explicit license to use and detailed source information. Reusable data must retain its original richness in order to be replicable or able to be combined in different settings. An accessible and clear data usage license and the source information about the original creation of data are required. In addition, discipline-specific data and metadata standards should be used to provide rich contextual information allowing reuse.

2.3. Resource Allocation

The responsibilities listed in Table 1 will guide the monitoring and updating of the NANCY's DMP.

Table 1: Responsibility & Role Allocation of DMP

Partner	Roles	Responsibilities
SID	Data Management Plan Management	• Leadership of Task 1.4 • In charge of the DMP execution and administration • In charge of ensuring the adequacy and authenticity of the dataset • In charge of keeping track of the datasets with open access • In charge of creating and maintaining the Dataset template included in Annex A.
DRAXIS	Dissemination Management	• In charge of overseeing the scientific publications, ensuring their availability through open access
OTE	Exploitation and Innovation Management	• Responsible for the further utilisation of DMP outcomes for research and innovation activities

[7] For further informative reading also see, *inter-alia*: Guide to Open Access in Horizon Europe, accessible at: https://enspire.science/guide-to-open-access-in-horizon-europe/

UOWM	Project Management	• Responsible for collaborating with the DMP manager to ensure timely updates of the DMP in the periodic reports.

All costs related to ensuring that publications with open access are included in the EU Grant money for the Horizon Europe (HE) project and that data produced by NANCY comply with the FAIR principals. The NANCY's project budget includes the costs necessary to make the data gathered or created during project operations FAIR. These costs encompass a specific set of data administration and processing duties. The following data manipulation and administration actions are taken into account:

- Collection
- Documentation
- Storage
- Access & Security
- Preservation
- Availability & Reuse
- Overall Data Management

Task 1.4's leading beneficiary is in charge of managing the data gathering in accordance with this data management strategy. The costs associated with storing data on NANCY's internal storage and making data accessible after the project is complete are included in the project's indirect costs. Each NANCY participant (universities, multi-actors, and SME partners) will be tasked with maintaining their own data in compliance with the DMP. This responsibility extends to individual researchers as well as teams of researchers. NANCY partners and researchers shall share the NANCY DMP with non-NANCY coworkers (if any) while ensuring that essential research data is kept in line with the relevant HE standards and methodologies. In this instance, this conformance shall be the exclusive responsibility of the original author/publisher/creator.

2.4. Security of Data

The NANCY datasets will be accessible to some or all of the partners and may be rendered publicly available or kept private. In both situations, the security of information is of the utmost significance. There are two distinct ways to explain the steps undertaken to guarantee the privacy of the data.

- A security measure is something that is used to safeguard private information while it is being shared or kept [8]
- Individual data is secured, in compliance with the General Data Protection Regulation (GDPR) [9]

Members of the NANCY consortium, adhere to the high standards of behaviour in the aforementioned topics and hold privacy and data protection in the greatest esteem. Security of information, which refers to how the data is moved and maintained, is influenced by the degree of protection of the data. To that end, for NANCY datasets, the subsequent rules shall be followed to guarantee the security of data:

Table 2: Data Security Scaling & Options

Type	Guideline
Confidential data of a partner	Deposit information on their own personal data storage
Confidential data between two partners	Deposit information on their own personal data storage or deposit information in a mutually collaborative environment

Confidential data between all partners	Deposit information on NANCY's internal storage
Publicly available data	Deposit and share information through Zenodo

The following recommendations have to be followed as a preventative measure:

- When utilising local storage, store information in a minimum of two distinct sites to prevent data loss.
- When utilizing cloud storage, check whether a data policy exists, guarding against unintentional deletions, like a recycling bin.
- Guarantee the overall dataset's uniformity by systematically assigning labels to files.

NANCY can be considered a very scientific-oriented project, having set high standards of scientific accomplishments throughout the project's life cycle. NANCY partners and researchers should be sure that their conclusions are supported by solid evidence, and external parties should have the capacity to confirm their scientific credibility by reviewing datasets and archives.

The information that follows will be recorded:

Information utilised in publications

- For instance, the data-gathering methods may include alternative items like surveys, photographs, videos, and other relevant materials.
- For instance, a database containing the research data that has been studied (in the form of worksheets and periodic summaries), in addition to the supporting documentation (such as the date the data was gathered, where it was collected from, and other relevant information).

Information needed by the engaged NANCY partners

- Recording of operations (protocol papers, requests for regulatory clearances, risk evaluations, and other relevant material).
- Documents relating to the financial administration of the project (PDF invoices, any additional accompanying financial documents, and other relevant material)

Information eligible for future scientific work

- The types of information that might be used in upcoming studies include intermediate reports, conclusion reports, and manuscript drafts in the literature.

The following categories of data will be retained, according to any restrictions enforced by GDPR and ethical authorisation processes (such as participant anonymity):

- Authentic information, such as audio and visual recordings, papers or digital records, hard copy surveys, and any other relevant material
- Unprocessed information, such as unprocessed digital data, logs of network activity, unprocessed pictures and/or videos, and other relevant materials
- Data sets that have been processed and employed in publications, together with information that helps individuals comprehend the processing (such as coding stages, data cleaning logs, missing values, and metadata for each dataset)
- In certain cases, secondary data sets are used to ensure the scientific accuracy and transparency of the data processing.
- Data sharing agreements

3. Data Management Plan monitoring & update

The DMP is a critical component of the project that involves the collection, storage, and utilisation of data. To ensure the accuracy and relevance of the DMP, it is essential to update it whenever substantial changes may occur. These changes could include the availability of new datasets that were not initially considered, the acquisition of new information that impacts data handling processes, modifications in the consortium's policy regarding data management, alterations in the makeup of the consortium itself, or even external factors that necessitate adjustments.

Regular updates to the DMP serve as a means of maintaining its integrity and effectiveness. By incorporating new datasets, partners can enhance the scope and depth of their analyses, leading to more comprehensive and accurate findings. Additionally, new information that becomes available during the lifespan of the project may require changes to the data management strategies outlined in the original plan. This ensures that the project remains aligned with the latest insights and developments in the field. The DMP will be updated during the project whenever substantial changes occur, such as new datasets, new information, changes in the consortium's policy, changes in the consortium's makeup, or external causes. The updates will be reported in the project's periodic reports.

Changes in the consortium's policy or the makeup of the consortium itself can also have a significant impact on data management practices. If the consortium decides to adopt new policies or guidelines related to data handling, it becomes crucial to update the DMP accordingly. Similarly, any potential modifications in the consortium's composition, such as the addition or departure of partners, may necessitate adjustments to the data management processes outlined in the plan.

To ensure transparency and accountability, all updates made to the DMP should be documented and reported in the project's periodic reports. This practice allows project partners to stay informed about the evolving data management strategies and any factors that might affect the integrity of the project's data. By maintaining a clear and comprehensive record of DMP updates, partners can demonstrate their commitment to rigorous data management practices and ensure the reliability of their findings.

In conclusion, updating the DMP throughout the project is essential to accommodate substantial changes such as new datasets, new information, policy modifications, consortium makeup alterations, or external causes. These updates enable partners to harness the potential of additional data, adapt to evolving circumstances, align with consortium policies, and maintain transparency. Regular reporting of DMP updates in the project's periodic reports enhances accountability and also ensures that stakeholders are properly aware of the changes made to data management strategies. By diligently updating the DMP, partners can uphold the highest standards of data management and enhance the overall quality of their project.

4. NANCY Datasets Overview

This section includes a high-level elaboration of the datasets that will be used or generated in the NANCY project. NANCY's internal repository (i.e., Confluence) will be utilised to store the project's dataset information. By utilising Confluence, the project team can ensure the continuous monitoring of the dataset's status, keeping track of any changes, updates, or revisions that occur throughout the project's life cycle. This allows for greater transparency and accessibility to the dataset information, facilitating collaboration and effective decision-making within the project.

Below is a summary table of the datasets utilised or created within the NANCY research project, as identified during the initial six months of the project. The details about each of these datasets are provided in Table 3. Furthermore, Annex B includes the completed templates corresponding to each of those datasets. The datasets can be changed and improved during the course of the project. The NANCY internal repository will have the latest version accessible there.

Table 3: NANCY Dataset Description

Partner Name	Dataset Name	Dataset Description	Datatype / format
UMU	ToN-IoT	The ToN-IoT datasets are new generations of Industry 4.0/Internet of Things (IoT) and Industrial IoT (IIoT) datasets for evaluating the fidelity and efficiency of different cybersecurity applications based on Artificial Intelligence (AI), i.e., Machine/Deep Learning (ML/DL) algorithms.	CSV files (processed) and raw traffic traces
UMU	UMU-5G	The dataset is composed of flow statistics calculated from real traffic, containing both benign and malicious samples, reproducing some kinds of attacks over control and user plane of the 5G testbed deployed in UMU	CSV file
UMU	5G-NIDD	5G-NIDD contains data extracted from a 5G testbed. The testbed is attached to 5G Test Network (TN) in University of Oulu, Finland. The data are extracted from tow base stations, each having an attacker node, and several benign 5G users. The attacker nodes attack the server deployed in the 5GTN MEC environment. The attack scenarios include DoS attacks and port scans. Under DoS attacks, the dataset contains ICMP Flood, UDP Flood, SYN Flood, HTTP Flood, and Slowrate DoS. Under port scans, the dataset contains SYN Scan, TCP Connect Scan, and UDP Scan.	CSV files (processed) and raw traffic traces
Bi2S (analysis / storage responsible)	Service performance	The dataset is composed of information regarding the performance of various services and	CSV file

		NFVs. The performance is measured in terms of computing units, storage units, latency, and bandwidth consumed.	
Bi2S (analysis/storage responsible)	UE/MEC node resource utilisation	The dataset is composed of data related to resource utilization rates (per UE and per MEC node) and number of tasks under processing (per UE and per MEC node). The term "resources" refers to the available computation and storage resources being utilised by the corresponding device. The data are collected throughout large periods of time with varying network traffic.	CSV file

5. NANCY IPR management

5.1. Introduction

Intellectual property rights (IPR) are linked with the legal protections of the outcomes of intellectual endeavours in almost all the sectors such as science, business, literature, and arts. IPR can be treated as a credible and clear way to participate in the development and rollout of cutting-edge 5G tools and technologies. This option is most relevant to suppliers and operators who participate in a project and hope to take advantage of its results.

Intellectual Property (IP)[8] is often categorised into two primary divisions[9]: (i) copyright and relevant rights, which cover broadcasts and performances, as well as scientific and artistic works, and; (ii) industrial property, which includes geographical indications, industrial designs, and invention patents. In research and innovation, patents[10] are the most commonly employed type of IPR. To qualify for patent protection, an invention must present practical utility and a level of novelty that is not included in the body of knowledge of the relevant technical field (what lawyers call the prior art). However, these requirements of practicality and novelty are not always enough. The invention should also involve a step towards an advancement – something that goes beyond the existing knowledge of the relevant field and that could be not deduced by someone with average knowledge of it. Patent holders possess the sole right to engage in commercial activities involving the sale, distribution, importation, and utilization of their patented innovations within the jurisdiction specified in the patent for the duration of its protection. The concepts of patent laws and IPRs can be challenging for researchers, entrepreneurs, and SMEs to navigate and understand.

IPR protection strategies take into account conditions of use and exploitation, IP protection and maintenance, IP monitoring and infringement, governing law, jurisdictions, or alternative dispute resolution (ADR) systems.

IPR management plays an important role in every data management program[11]. A database, or other data resource, creator will be probably interested in who owns that resource and how others should use it. Any individual, or party, that may populate that resource with data, which partially origins from third parties will want to make sure that all legal, ethical, and professional obligations that one may have to the provider of the data are met. Since the benefits of data sharing are so well known and

[8] Intellectual property (IP) management is a key element in improving the competitiveness of any legal entity. In the context of ongoing EU-funded projects, recommendations and guidelines for the management of IPs are also provided by the European Commission; for example, the document: IPR Helpdesk (2014): Fact Sheet: *How to manage IP in Horizon 2020: at the implementation stage.* European Union. Available at: https://www.uv.es/operuv/docs_h2020/IP_Management_h2020_implementation_0.pdf . In particular, this document covers issues about: (i) Contractual agreements among project beneficiaries for joint and individual exploitation of the IPs, including the use of background information; (ii) directions for the exploitation and dissemination of the outcomes, and; (iii) directions for the protection of IP, including technologies, software, and approaches for commercial exploitation.

[9] World Intellectual Property Organization (WIPO) (2020): *What is Intellectual Property?* (WIPO Publication No. 450E/20). Available at: https://tind.wipo.int/record/28588

[10] A patent is an exclusive right granted for the protection of inventions (products or processes) offering a new technical solution or facilitating a new way of doing something. The patent holder enjoys the exclusive right to prevent third parties from commercially exploiting their invention for a limited period. In return, the patent holder must disclose the invention to the public in the patent application.
Also see, *among others,* the detailed scope discussed in the document: European IP Helpdesk (2019): *Your Guide to IP in Europe*. European Union. Available at: https://op.europa.eu/en/publication-detail/-/publication/e8312ba4-4aa1-11ed-92ed-01aa75ed71a1/language-en/format-PDF/source-273095800

[11] For further information also see: https://data.research.cornell.edu/content/intellectual-property

documented, a researcher may wish to share their database and/or content with others, while others can only fully utilise external data if they are aware of the terms of use (if any) for that data.

NANCY is a collaborative effort involving synergetic actions between participants from many different countries and market sectors. This purely implies that a well-defined strategy, which will set the guidelines for the "appropriate handle" of the data management and the IP ownership – where relevant –, is a process of critical importance to preserve, secure, protect, and evaluate/improve any generated value, related impact, or, in general, project's results, especially for those cases related to the use of software tools.

On the other hand, by taking into account the fact that according to the global market-oriented principles, contract formalisation of software follows the format of license agreements, which impose specific usage rules for third parties that intend to utilise the related (software-based, or other) outcomes. The development of specific software can be assessed as a purely important action, as it is most likely to be dependent on the coding and development abilities of a developer/creator as well as the ability to suitably and appropriately convert various functional and/or operational requirements that have been identified in the process of the designing phase into a sequence of instructions.

Figure 3 illustrates a typical software development process and the various IPRs that can be generated.

A software process model[12] is an abstraction of the software development process. This model specifies the stages and the order of the processes. So, it is a sort of representation of the order of activities of the process and the sequence in which they have to be performed. Thus, such a model can define (i) the tasks to be performed; (ii) the input and output of each task; (iii) the pre- and post-conditions for each task, and; (iv) the flow and sequence of each task.

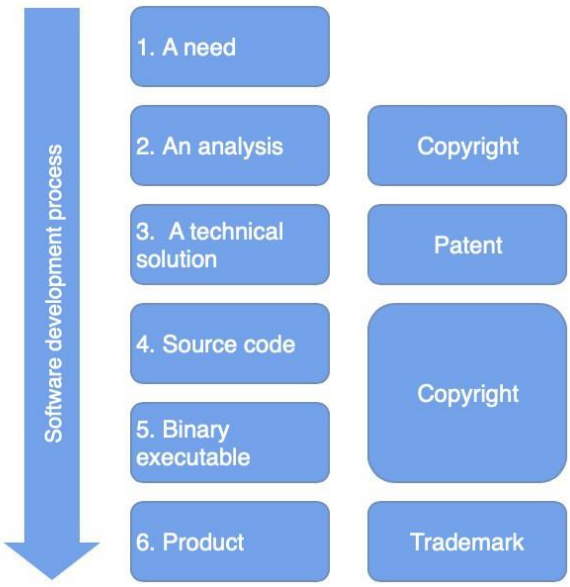

Figure 3: Depiction of a software development process

Various research data generated or created in the context of a project framework may include, *for example*, statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, and images.

[12] More details can be found, *for example*, at: https://www.educative.io/blog/software-process-model-types#what. Also see: https://cynoteck.com/blog-post/top-software-development-models-to-choose-from/

In this point, it would be wise and essential to clarify that contractual obligations to disseminate results (as defined in Article 17 of the GA) and to provide open access[13] to scientific publications do not, in any way, change the obligation of consortia to protect project's results, ensure the confidentiality and security obligations, or the obligations to protect personal data, all of which still apply.

In the successful exploitation of any project, it is of high importance that all beneficiaries have reach a consensus concerning the ownership of IP, the results access rights, and the background IP for the project execution, as well as the protection of confidential information

Within NANCY's GA, IPR-related issues (including background and results, access rights, and rights of use) as well as communication, dissemination and visibility issues are covered within Articles 16 and 17, respectively.

5.2. IPR Management Process

In principle, the corresponding project Consortium Agreement (CA) sets and clarifies the essential guidelines that manage and control the access rights[14] of the different IPRs, both for the contributed background and for the foreground or results[15]. Normally, such guidelines describe the rights that anyone of the involved actors have regarding the IPRs that are specific to the project set of activities. Regarding involved "actors", they can be either the project partners or external parties. In what concerns the project partners, they have been granted free access to the foreground[16] produced by any other partner in the consortium throughout the project lifetime and after, explicitly for the purposes of achieving interoperability of software, for commercial and non-commercial research, for dissemination[17] and for commercial and/or non-commercial demonstration, while for other potential purposes, they can have access under fair and reasonable conditions[18]. On the other hand, access rights to background[19], owned by other project partners, are to be granted under fair and reasonable conditions that need to be negotiated in good faith between the concerned parties.

[13] "Open-access" implicates for online access to research outputs provided free of charge to the end-user. NANCY will make certain to produce open-access research; specifically, the project aims in specifying the research plans to be followed as early as possible and making the aforementioned plans publicly available through preregistration. To secure high quality research outputs, scientific publications under development will be subjected to peer-reviews prior to the data collection process, as registered reports. NANCY aims in enriching public knowledge and support scientific evolution, through open access to insightful research results and technological advancements.

[14] The term implicates for any "rights to use results or background".

[15] According to the NANCY Grant Agreement (GA), as in Article 16§2, the term "results" means any tangible or intangible effect of the action, such as data, know-how, or information, whatever its form or nature, whether or not it can be protected, as well as any rights attached to it, including intellectual property rights

[16] Normally, the foreground knowledge will be the ownership of involved partners. Any foreground knowledge that could be jointly created by at least two project partners is regulated by the context of the dedicated Consortium Agreement. The consortium can enter into an agreement regarding the ownership of foreground and access rights to be provided to any partner for use and dissemination purposes. Any details concerning the exposure to jointly owned foreground knowledge, joint inventions, and joint patent applications are "addressed" in the CA.

[17] "Dissemination" means the public disclosure of the results by appropriate means, other than resulting from protecting or exploiting the results, including by scientific publications in any medium.

[18] "Fair and reasonable conditions" implicates for appropriate conditions, including possible financial terms or royalty-free conditions, taking into account the specific circumstances of the request for access, for example the actual or potential value of the results or background to which access is requested and/or the scope, duration or other characteristics of the exploitation envisaged.

[19] According to the NANCY Grant Agreement (GA), as in Article 16§1, "background" is defined as any data, know-how or information – whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights – that is: (i) held by the beneficiaries before they acceded to the GA, and; (ii) needed to implement the action or exploit the results. If background is subject to rights of a third party, the beneficiary concerned must ensure that it is able to comply with its obligations under the NANCY GA. In principle, project beneficiaries must identify in a written agreement the background as needed for implementing the related project action or for exploiting its results.

For the cases where external parties may be involved, each project partner which had developed a component/software – or part of it – should take the decision on whether to distribute or make available the code of its component/software under controlled and proprietary license terms or as open-source[20].

The role of the IPR management process is to "delineate" a strategy aiming to protect the project's results and, simultaneously, to steer the exploitation roadmap of the partners. Such sort of process can be made of four steps and has to remain active for the entire lifecycle of the duration of the project. It comprises the following "steps":

1. Background identification: The first step starts during the proposal preparation phase when several – or all – of the partners have listed their know-how and/or software that forms the background of the project.

2. Sideground identification (optionally): As with the previous step, the identification of the possible project sideground is continuously updated during the project life span.

3. Foreground identification: All the partners who are part of the development of software components – or machine learning algorithms and models– are requested to list and make them available on a dedicated project code repository. This repository is private and accessible only to project partners by means of credentials.

4. Foreground classification: Each partner who develops software components or machine learning algorithms/models is in charge of deciding which license to be used for these components, according to the license listed and described in a dedicated part within the IPR management framework.

5. Rules and policies: All project partners are requested to follow the policies and rules stated in the CA and the GA.

Background

All NANCY partners brought some assets to the project such as data, know-how, or software, that have been identified and documented during the proposal preparation phase and listed in the respective section of the CA. These assets constitute the project's background. The NANCY partners have identified the background needed for the project to avoid this use (or re-use) of background may lead to any breach of obligations or, possibly, an infringement of intellectual property rights belonging to others.

Sideground

Partners may be able to report any sideground, corresponding to software and knowledge generated by the partners during the project period but outside the "strict" contractual project effort.

Foreground: Assets identification

Based on the progress of the project, as it is pictured in dedicated project deliverables conformant to the GA provisions (i.e. in dissemination and exploitation reports), the partners should be able to decide to list all the components – or related software modules – that concise the explicit results, outcomes, or findings of NANCY R&D activities. If so, for each of such components it is necessary to specify the corresponding ownerships among the project beneficiaries, the related project WP(s), and the associated license type. Most of the foreground knowledge is associated with well-known open-source license types, while some of them can be used and distributed under proprietary licenses.

[20] "Open source" is a term that originally referred to open source software (OSS). Open source software is code that is designed to be publicly accessible; so anyone can see, modify, and distribute the code as they see fit. The Open Source Initiative (OSI) is a non-profit corporation with global scope formed to educate about and advocate for the benefits of open source and to build bridges among different constituencies in the open source community. For further information see: https://opensource.org/

5.3. Intellectual Property Rights Plan

The NANCY consortium adheres to the European Commission's directions (as in the IPR Helpdesk Fact Sheet) to address issues concerned with the management of IP and effectively protect the IPR in the most efficient way. To this aim, an IPR plan has been established, as shown in Figure 4.

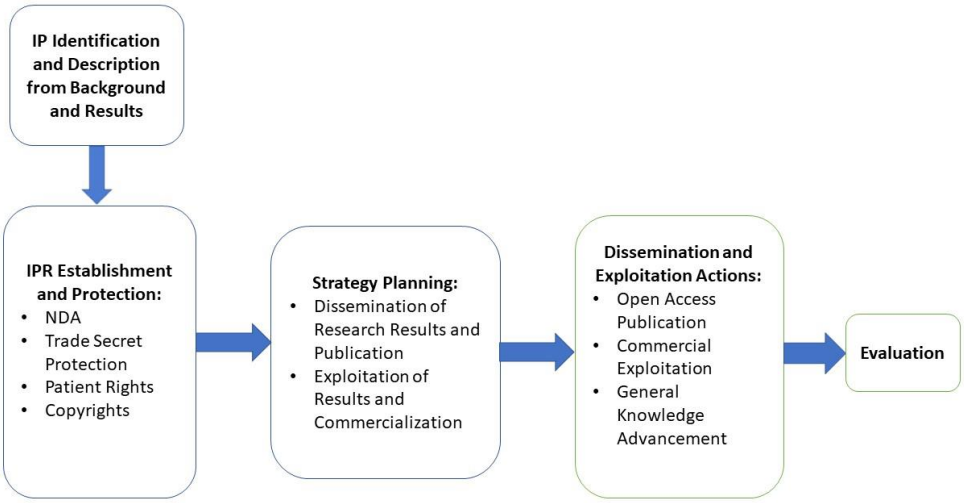

Figure 4: Detailed actions for an IPR Plan

In the following, we briefly present the conceptual framework of each one of the above "steps":

1. Identification of IP from background and results: The background of the NANCY project has been identified since the beginning of the project and has become binding via the CA provisions. Throughout the project implementation activities, NANCY's partners/beneficiaries will identify the project results that will be generated (initial estimated project results are already available in the Description of Action (DoA) of the GA). All potentially exploitable IPs will be identified and recorded accordingly. Of note, the IP-related outcomes can originate from one or multiple NANCY components.

2. **Description of IP interfaces:** This step is focused on drafting a clear description of the (corresponding) products and their interfaces with other NANCY's components, requirements, etc. to identify potential extensions of them.

3. **Establishing ownership of IPR:** After identifying and describing the corresponding product(s) based on NANCY's results/outcomes, the consortium will identify the ownership and/or joint ownership of the respective results/outcomes. For this purpose, an IPR identification sheet can be utilised.

4. **IPR protection measures:** The Exploitation and Innovation Manager will monitor the generated knowledge on a regular basis in collaboration with the involved beneficiaries. Non-Disclosure Agreements (NDAs) can be utilised on a peer-to-peer basis during the intended collaborative effort. If required, the NANCY consortium may ask for legal advice for identifying and assessing patent application or copyright declaration opportunities. Additional IPR protection measures can include a description of innovation elements, by reviewing similar innovations and/or existing patents, proposing patent applications, and reporting the status to the consortium. The IPR protection measures for NANCY will be addressed both on beneficiary an on consortium levels.

5. **Dissemination and exploitation strategy:** In this step, the consortium will draft potential strategies for the dissemination and exploitation of IPRs, such as open access and commercialisation of products.

6. **Dissemination and exploitation actions:** This step consists of various means of protecting the dissemination and exploitation. Such means include patent filing, commercialization of products, publications in scientific journals, etc.

7. **Evaluation:** The final step is focused on the assessment and evaluation of the effectiveness of the IPR protection measures, dissemination, and exploitation activities.

5.4. NANCY's Specific IPR Management Framework

In the context of NANCY, IPR will be handled according to the general European Commission's policies regarding ownership, exploitation rights, and confidentiality. Furthermore, project participants are encouraged to use the European IPR Helpdesk[21] for questions related to the protection of IPR. A Consortium Agreement (CA) has already been signed by the partners since the project kick-off meeting, governing a variety of activities, in which IPR are present too. The CA document has been based on the latest version of the MCARD-HEU Model template provided by DigitalEurope[22], allowing all participants to be able to take advantage of exploitation opportunities, with respect to the IPR and individual exploitation interests.

As a general policy, the IP Rights for common, and new developments within the project will remain with the original author. IPR management within the consortium was critically relevant such that key existing patents ("background information") had to be identified before the project started. IPR ownership is assessed within the CA.

Within NANCY's context, IPR means patents, patent applications, and other statutory rights in inventions; copyrights (including without limitation copyrights in software); registered design rights, applications for registered design rights, unregistered design rights, other statutory rights in designs; and other similar or equivalent forms of statutory protection, wherever in the world arising or available.

Results shall have the meaning given to them in the GA, meaning any tangible or intangible effect of the project, such as data, knowledge, and information regardless of their form or nature, whether or not they can be protected, which are generated in the project as well as any rights attached to them, including IPR. Results do not take into account the effects generated/produced by activities outside of the NANCY project – be it before the project start, during its course, or after it ends.

The main lines to be followed regarding IPR management during NANCY's course are listed below:

- ***Licensing of pre-existing know-how (Background knowledge):*** The introduced background knowledge will remain property of each partner, meaning that it has been already agreed that, during the project, background knowledge will be made available to participants in the project in such a way in order not to undermine the full usability (at no cost) for the consortium of related foreground knowledge.
- ***Knowledge gained within the project (Foreground knowledge):*** The foreground knowledge will be owned by the involved partners. Any foreground knowledge that could and was collaboratively/jointly created by at least two project partners is regulated by the Consortium

[21] The European IP Helpdesk supports European SMEs and research teams involved in cross-border business and/or EU-funded research activities manage, disseminate, and valorize their IP. Offering a broad range of informative material, a Helpline service for direct IP support as well as on-site and online training, our main goal is to support IP capacity building along the full scale of IP practices: from awareness to strategic use and successful exploitation. More details about this can be found at: https://intellectual-property-helpdesk.ec.europa.eu/regional-helpdesks/european-ip-helpdesk_en

[22] For more informative details also see: https://www.digitaleurope.org/resources/digitaleurope-mcard-heu-model-consortium-agreement-for-research-development-and-innovation-actions-under-horizon-europe/

Agreement. The consortium will enter into an agreement regarding the ownership of foreground and access rights to be provided to any partner for use and dissemination purposes. CA addresses any details concerning the exposure to jointly owned foreground knowledge, joint inventions, and joint patent applications.

According to the CA, a procedure has been established to **inform** all other partners in the project when the results of the project will be made public (e.g., in a conference paper), in order to allow partners to protect IPR before publication. NANCY aims to have **patents** filed by one – or more partners – with the main goal to protect innovations from the project which have the potential to have a significant impact. Innovations, concepts, and solutions that will not be protected by patent applications by the participants will be made public after an agreement between the partners and the consortium, where relevant. Procedures established in the CA will be followed to prevent others from blocking the usage of these results.

NANCY will strongly **advocate open-source licenses**[23] for SW releases. It is intended that most of the project SW prototypes, especially the ones developed by academic partners, will be based on open-source software and will be open-source. In the case of SW prototypes that have been jointly developed by two or more project partners, the respective SW will be released under open source only in the case that all the involved partners unanimously agree. Regarding project public deliverables, the partners of the project will jointly agree on the licenses under which the deliverable will be made available, separately for each deliverable. If an agreement cannot be reached, then the respective deliverable(s) will be considered proprietary.

IPR will be managed according to principles of equality of all the partners towards the foreground knowledge and in full compliance with the general European Commission policies, particularly regarding: (i) ownership; (ii) exploitation rights, and (iii) confidentiality. In general, outcomes, innovative ideas, concepts, and solutions where protection by patent applications is not pursued by the partners and which are not characterised as trade secrets will be made public after and if an agreement between the partners can be made, to allow the community to benefit from these results and exploit them. The CA is used as a reference for all IPR cases simultaneously.

The Consortium Agreement provides guidelines and rules for handling confidentiality and IPR to benefit the consortium and its partners. All the project documentation will be electronically stored in the project's collaborative environment which is maintained by the coordinator. Classified documents will be handled according to the respective rules with regard to classification, numbering, locked storage, and distribution limitations. The policy that will govern the IPR management in the scope of NANCY's activities, is practically based on the principles described in the CA. IPRs mentioned in the CA are:

- Copyrights (including, but not limited to, copyrights in software applications)
- Patents, patent applications, and other legal privileges in inventions
- registered creation privileges, applications for registered creation privileges, unregistered creation privileges, and other lawful rights in designs
- additional similar or equivalent forms of legal protection, emerging or available anywhere in the world

[23] Open source licenses are licenses that comply with the Open Source Definition; in brief, they allow software to be freely used, modified, and shared. To be approved by the Open Source Initiative (also known as the OSI) a license must go through the Open Source Initiative's license review process. Popular open source licenses include the Apache License, the MIT License, the GNU General Public License (GPL), the BSD Licenses, the GNU Lesser General Public License (LGPL) and the Mozilla Public License (MPL). For further details also see, *inter-alia*: https://opensource.org/licenses/

6. Ethical Aspects

A component of contemporary research ethics is data security and privacy. When organising NANCY's project operations, there are a number of difficulties that must be properly dealt with and handled. Compliance with the legal and ethical guidelines is an essential basis for the work of a modern researcher, so properly defining the legal and ethical concerns (such as GDPR, meticulous awareness of sensitive data, and minimising the burden placed on individuals engaged in research) facilitates more efficient data handling now and in the future.

Ethics flows through all aspects of research, and ethical adherence is essential to achieving true scientific excellence for all activities conducted under the NANCY project. The application of fundamental ethical laws and guidelines to scientific research across all potential fields of study is implied by ethical research conduct. The Ethics Appraisal Procedure [10] is the method used to evaluate, as well as manage the ethical aspect of activities supported by Horizon 2020 and Horizon Europe.

In the case that personal data or any kind of PII is gathered from the Use Cases (UCs), it will be safeguarded in compliance with GDPR [3], which relates to the open transmission of such data and to the confidentiality of natural people with regard to the storage and use of sensitive information. Should any of the gathered data include sensitive or private information, the consortium has agreed on the Consortium Agreement (CA), which includes a particular non-disclosure clause across all partners, to ensure that no data will be published or employed for any purpose other than the NANCY project.

According to the NANCY's CA, "Sensitive Information" is considered any information shared by one party (the "Disclosing Party") to another party (the "Recipient") during the implementation of an action, which is explicitly marked as "confidential" or "secret" at the time of disclosure, or identified as sensitive when shared verbally and confirmed in writing within 15 days. If a stakeholder must reveal "Sensitive Information", he/she shall notify the Disclosing Party in advance and, to the extent permitted by law, shall follow the Disclosing Party's reasonable instructions to protect confidentiality. Each recipient must inform the disclosing party in writing as soon as they become aware of any unauthorized disclosure, misappropriation, or misuse of sensitive information.

According to Horizon's 2020 "Regulation of Establishment" for ethical principles (Art. 19) [11], all research and innovation projects must adhere to moral standards and pertinent national, international, and European Union laws, including the European Union Fundamental Rights Charter, European Convention on Human Rights, and its Additional Protocols. The right to confidentiality, the right to safeguard private information, the need to provide high standards of human health protection, and the concept of proportionality are all things that should get special consideration. Explicit scientific domains that are not eligible for financial support, are the following:

- Research conducted with the goal of cloning humans for reproductive purposes.
- Research aimed at altering the genetic makeup of human beings in a way that can be passed down to future generations.
- Research carried out to generate human embryos exclusively for research purposes or for obtaining stem cells, including through a process called somatic cell nuclear transfer

The legislative framework of the participating Member States as well as the content of the scientific proposal will determine whether funding for research on adult and embryonic human stem cells is

possible. In any case, NANCY's activities and implementations as described in the DoA, do not include any such endeavours and the aforementioned concepts are not applicable for NANCY.

NANCY has implemented the notion of "privacy by design," which offers an approach for centering the architecture of systems, databases, and procedures on respect for data subjects' basic rights, in order to develop ethically and responsibly. The GDPR, which NANCY complies with, now incorporates the broader idea of "data protection by design" [12], which calls on data controllers to take the proper organisational and technological steps to put its fundamental data-protection principles [13] into practise (see paragraphs 5 and 25 of GDPR). One of the greatest methods to handle the ethical issues raised in the project's life cycle is to include data protection by design. Some key actions to accomplish data protection by design could be:

- Pseudonymization or anonymization of personal data
- Minimizing the collection and storage of data (data minimization)
- Utilizing cryptography techniques such as encryption and hashing
- Engaging data protection-focused service providers and storage platforms
- Establishing mechanisms that allow individuals to exercise their fundamental rights, such as accessing their personal data directly and giving consent for its use or transfer.

7. Conclusion

This deliverable is focused on the NANCY data and IPRs. In-depth descriptions of the NANCY DMP and IPR management strategies are included as components of the present work. In both instances, the EU guidelines and legislation, as well as the CA document have been taken into consideration to formulate these strategies and approaches.

The NANCY DMP and IPR administrative plans, which both make full use of the NANCY collaboration platform, empower NANCY's partners to track, supervise, and modify digital information and IP assets as necessary.

Five NANCY datasets were defined in the first six months of the project. Using the given templates, NANCY stakeholders can suggest new datasets and IP assets or change current ones. The status of the NANCY datasets and intellectual property will be reported during the forthcoming plenary meetings and will be assessed in the management reports as contractually defined.

Bibliography

[1] "Template Horizon 2020 Data Management Plan (DMP)," [Online]. Available: https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf

[2] R.-D. Veit, "Safeguarding Regional Data Protection Rights on the Global Internet—The European Approach Under the GDPR," in Personality and Data Protection Rights on the Internet, Springer, Cham, vol. 96, pp. 445-484, Mar. 2022.

[3] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation," [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[4] "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/LSU/?uri=celex:31995L0046

[5] European Comission, "EU support for Open Access." [Online]. Available: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-access_en

[6] Zenodo, "ZENODO General Policies v1.0." [Online]. Available: https://about.zenodo.org/policies/

[7] OpenAIRE, "Licenses for Research Data." [Online]. Available: https://www.openaire.eu/how-do-i-license-my-research-data

[8] I. Gupta and A. K. Singh, "A Holistic View on Data Protection for Sharing, Communicating, and Computing Environments: Taxonomy and Future Directions," arXiv:2202.11965 [cs.CR], 2022. [Online]. Available: https://arxiv.org/abs/2202.11965

[9] J. Trzaskowski and M. G. Sorensen, GDPR Compliance: Understanding the General Data Protection Regulation, Ex Tuto Publishing A/S, 2019.

[10] European Commission, "Ethics." [Online]. Available: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm

[11] "Horizon 2020 - Regulation of Establishment: Ethical principles (Article 19)," [Online]. Available: https://ec.europa.eu/research/participants/data/ref/h2020/legal_basis/fp/h2020-eu-establact_en.pdf#page=11.

[12] Intersoft Consulting, "Art. 25 GDPR - "Data protection by design and by default." [Online]. Available: https://gdpr-info.eu/art-25-gdpr/

[13] Intersoft Consulting, "Art. 5 - "GDPR Principles relating to processing of personal data." [Online]. Available: https://gdpr-info.eu/art-5-gdpr/

Annex A

The template for the partner's datasets summary can be found in the Table 4. The table includes instructions regarding how to properly compile it. This template will be used throughout the project's life cycle to facilitate the dataset documentation. The corresponding updates and enhancements will occur at M12 and M24 internally, as well as at M18 and M36 at the first and second official periodic reports of the project. Such changes could be the inclusion of new datasets, availability of new information, modifications in the consortium's policies, alterations in the consortium's composition, or external factors that may impact the project.

All the below sections and sub-sections are provided with the intention to capture all relevant and necessary information from the datasets that are provided in terms of the NANCY project from the involved beneficiaries.

The primary sections which constitute this template are the following:

- Dataset Summary
- NANCY Work Packages, Tasks and Deliverables
- Partners Services and Responsibilities
- FAIR Data – Making data findable, including provisions
- FAIR Data – Making data openly accessible
- FAIR Data – Making interoperable
- FAIR Data – Increase data re-use
- Allocation of Resources
- Data Security
- Ethical Aspects

Table 4: NANCY Dataset Template with Explanations

Dataset Name	Please provide the name of the Dataset
Dataset Summary	
Dataset Description	Please provide a Description for the Dataset.
Dataset Purpose	Please describe the purpose of the Dataset.
Dataset Type/Format	Please provide the type/format of the Dataset (e.g MySQL db, CSV file e.t.c).
Re-use of existing data	Please explain if this Dataset is expected to re-use previously collected data.
Dataset Origin	Please provide the Dataset origin.
Dataset Collection	Please describe how the Dataset was collected.
Expected Size	Please provide an estimation about the size of the Dataset.
NANCY Work Packages, Tasks and Deliverables	
Relevant Work Packages	Please provide the WP(s) where the Dataset will be utilized.
Relevant Tasks	Please provide the Task(s) where the Dataset will be utilized.
Relevant Deliverables	Please provide the Deliverable(s) where the Dataset will be utilized.
Partners Services and Responsibilities	
Partner Owner	Please provide the owner of the Dataset.
Partners responsible for data collecting	Please provide the partner who will be responsible to collect the Dataset.
Partners responsible for data analysis	Please provide the partner who will be responsible to analyse the Dataset.
Partners responsible for data storage	Please provide the partner who will be responsible to store the Dataset.
FAIR Data – Making data findable, including provisions	
Discoverable and Identifiable Data	Please describe if the data within the Dataset are expected to be discoverable and identifiable.
Naming Convention	Please provide a naming convention for the Dataset.
Versioning	Please provide if the Dataset will follow a versioning policy.
Search Keywords	Please provide if the Dataset is expected to support Search Keywords.
Metadata	Please describe if the Dataset is expected to keep metadata information.
FAIR Data – Making data openly accessible	
Dataset Openly Accessible	Please describe if the Dataset is expected to be openly accessible.
Methods/Tools for accessing the data	Please provide any methods for accessing the data.
Access Restrictions	Please provide any access restrictions.
Repository	Please provide the repository, if applicable.
FAIR Data – Making interoperable	
Interoperability	Please describe if the Dataset will be interoperable.
Standards	Please provide any interoperability standards.
FAIR Data – Increase data re-use	

Licence	Please provide any license which protects the Dataset.
Data sharing terms	Please provide any data sharing terms related to the Dataset.
Dataset sharing after the end of the project	Please provide any data sharing terms related to the Dataset, upon the conclusion of the project.
Preservation and update of the dataset after the project	Please provide any long-term support plan for the Dataset, upon the conclusion of the project.
Re-use timeframe	Please describe any re-usage timeframe.
Allocation of Resources	
Cost for making the dataset FAIR	Please provide the cost of creating this Dataset FAIR.
Costs for the preservation (including backup) and the update of the dataset	Please provide the costs of preservation and update of the Dataset.
Data Security	
Security Measures	Please provide any security measures for the Dataset.
Privacy Measure	Please provide any privacy measures for the Dataset.
Ethical Aspects	
Ethical and Legal Aspects	Please provide the Ethical and Legal Aspects of the Dataset
Consent form for data sharing and long preservation within the NANCY consortium	Please provide the consent form for data sharing and long preservation forms for the NANCY consortium.
Consent form for sharing the dataset publicly.	Please provide the consent form for the public data sharing of the Dataset.
Other Issues	Please provide any other issues.

Annex B

This section includes all the compiled templates of the datasets provided by the corresponding partners up until month six of the project, which will be used or generated in the NANCY project. Annex B will be enhanced with all the added responses from the partners during the internal, as well as the official update of the document.

Table 5: NANCY Dataset ToN-IoT

Dataset Name	ToN-IoT
Dataset Summary	
Dataset Description	The ToN-IoT datasets are new generations of Industry 4.0/Internet of Things (IoT) and Industrial IoT (IIoT) datasets for evaluating the fidelity and efficiency of different cybersecurity applications based on Artificial Intelligence (AI), i.e., Machine/Deep Learning algorithms.
Dataset Purpose	Anomaly/attack detection
Dataset Type/Format	CSV files (processed) and raw traffic traces
Re-use of existing data	No
Dataset Origin	Cyber Range and IoT Labs, the School of Engineering and Information technology (SEIT), UNSW Canberra @ the Australian Defence Force Academy (ADFA)
Dataset Collection	The datasets were collected from a realistic and large-scale network designed at the Cyber Range and IoT Labs, the School of Engineering and Information technology (SEIT), UNSW Canberra @ the Australian Defence Force Academy (ADFA).
Expected Size	Around 5 million samples
NANCY Work Packages, Tasks and Deliverables	
Relevant Work Packages	Please provide the WP(s) where the Dataset will be utilized.
Relevant Tasks	Please provide the Task(s) where the Dataset will be utilized.
Relevant Deliverables	Please provide the Deliverable(s) where the Dataset will be utilized.
Partners Services and Responsibilities	
Partner Owner	UMU
Partners responsible for data collecting	UMU
Partners responsible for data analysis	UMU
Partners responsible for data storage	UMU
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	Please describe if the data within the Dataset are expected to be discoverable and identifiable.
Naming Convention	Please provide a naming convention for the Dataset.
Versioning	No
Search Keywords	Yes
Metadata	Yes
FAIR Data - Making data openly accessible	

Dataset Openly Accessible	Yes
Methods/Tools for accessing the data	Download through a repository publicly accessible
Access Restrictions	No
Repository	https://cloudstor.aarnet.edu.au/plus/s/ds5zW91vdgjEj9i

Table 6: NANCY Dataset UNU-5G

Dataset Name	UMU-5G
Dataset Summary	
Dataset Description	The dataset is composed of flow statistics calculated from real traffic, containing both benign and malicious samples, reproducing some kinds of attacks over control and user plane of the 5G testbed deployed in UMU
Dataset Purpose	Anomaly/attack detection
Dataset Type/Format	CSV file
Re-use of existing data	No
Dataset Origin	UMU testbed
Dataset Collection	Deployment of monitoring probes over the testbed, benign traffic generation through simulations of different types of traffic and malign traffic generation based on both control plane attacks (e.g., 5GReplay) and user plane attacks (e.g., DoS/DDoS, Recon, etc.)
Expected Size	Around 200,000 samples
NANCY Work Packages, Tasks and Deliverables	
Relevant Work Packages	Please provide the WP(s) where the Dataset will be utilized.
Relevant Tasks	Please provide the Task(s) where the Dataset will be utilized.
Relevant Deliverables	Please provide the Deliverable(s) where the Dataset will be utilized.
Partners Services and Responsibilities	
Partner Owner	UMU
Partners responsible for data collecting	UMU
Partners responsible for data analysis	UMU
Partners responsible for data storage	UMU
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	Please describe if the data within the Dataset are expected to be discoverable and identifiable.
Naming Convention	Please provide a naming convention for the Dataset.
Versioning	Yes
Search Keywords	Yes
Metadata	Yes
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	Yes
Methods/Tools for accessing the data	Download through a repository hosted in UMU

Access Restrictions	NANCY partners
Repository	Yet not available.

Table 7: NANCY Dataset 5G-NIDD

Dataset Name	5G-NIDD
Dataset Summary	
Dataset Description	5G-NIDD contains data extracted from a 5G testbed. The testbed is attached to 5G Test Network in University of Oulu, Finland. The data are extracted from tow base stations, each having an attacker node, several benign 5G users. The attacker nodes attack the server deployed in 5GTN MEC environment. The attack scenarios include DoS attacks and port scans. Under DoS attacks, the dataset contains ICMP Flood, UDP Flood, SYN Flood, HTTP Flood, and Slowrate DoS. Under port scans, the dataset contains SYN Scan, TCP Connect Scan, and UDP Scan.
Dataset Purpose	Anomaly/attack detection
Dataset Type/Format	CSV files (processed) and raw traffic traces
Re-use of existing data	No
Dataset Origin	5G Test Network in University of Oulu
Dataset Collection	From OULU's 5G testbed's tow base stations
Expected Size	Around 1 million samples
NANCY Work Packages, Tasks and Deliverables	
Relevant Work Packages	Please provide the WP(s) where the Dataset will be utilized.
Relevant Tasks	Please provide the Task(s) where the Dataset will be utilized.
Relevant Deliverables	Please provide the Deliverable(s) where the Dataset will be utilized.
Partners Services and Responsibilities	
Partner Owner	UMU
Partners responsible for data collecting	UMU
Partners responsible for data analysis	UMU
Partners responsible for data storage	UMU
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	Please describe if the data within the Dataset are expected to be discoverable and identifiable.
Naming Convention	Please provide a naming convention for the Dataset.
Versioning	No
Search Keywords	Yes
Metadata	Yes
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	Yes
Methods/Tools for accessing the data	Download through a repository publicly accessible
Access Restrictions	No
Repository	https://etsin.fairdata.fi/dataset/9d13ef28-2ca7-44b0-9950-225359afac65/data

Table 8: NANCY Dataset Service Performance

Dataset Name	Service performance
Dataset Summary	
Dataset Description	The dataset is composed of information regarding the performance of various services and NFVs. The performance measured in terms of computing units, storage units, latency and bandwidth consumed.
Dataset Purpose	The dataset will be used to train ML models to optimise resource scheduling of services and NFVs in the B-RAN environment.
Dataset Type/Format	CSV file
Re-use of existing data	No
Dataset Origin	**Any** testbed that supports B-RAN
Dataset Collection	Deployment of monitoring software over several UEs/ MEC nodes (mobile devices, terminals or tablets). Software should measure the performance of different NFVs/services in terms of computing units, storage units, latency and bandwidth consumed.
Expected Size	Around 1 million samples
NANCY Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP4 and WP6
Relevant Tasks	T4.3, T6.5 – T6.9
Relevant Deliverables	D4.3, D6.5 – D6.9
Partners Services and Responsibilities	
Partner Owner	-
Partners responsible for data collecting	-
Partners responsible for data analysis	Bi2S
Partners responsible for data storage	Bi2S
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	YES
Naming Convention	Per service: Service name, computing units, storage units, latency, BW utilization.
Versioning	YES
Search Keywords	YES
Metadata	YES
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	YES
Methods/Tools for accessing the data	Download through a Zenodo repository
Access Restrictions	NONE
Repository	Not yet available
FAIR Data - Making interoperable	
Interoperability	Please describe if the Dataset will be interoperable.
Standards	Please provide any interoperability standards.

FAIR Data – Increase data re-use	
Licence	None required
Data sharing terms	Open data sharing
Dataset sharing after the end of the project	1 year after the project ends
Preservation and update of the dataset after the project	2 years after the project end
Re-use timeframe	Once within 6-month period
Allocation of Resources	
Cost for making the dataset FAIR	No cost is foreseen
Costs for the preservation (including backup) and the update of the dataset	For the backup: No costs are foreseen For updating the dataset: More experiments should be conducted in a NANCY testbed. Very small energy consumption costs are expected
Data Security	
Security Measures	None
Privacy Measure	None
Ethical Aspects	
Ethical and Legal Aspects	None
Consent form for data sharing and long preservation within the NANCY consortium	The consortium should sign a consent form in order to make the data open access.
Consent form for sharing the dataset publicly.	No consent form is required
Other Issues	No other issues are identified

Table 9: NANCY Dataset UE/MEC node resource utilisation

Dataset Name	UE/MEC node resource utilisation
Dataset Summary	
Dataset Description	The dataset is composed of data related with resource utilization rates (per UE and per MEC node) and number of tasks under processing (per UE and per MEC node). The term “resources” refers to the available computation and storage resources being utlised by the corresponding device. The data are collected throughout large periods of time with varying network traffic.
Dataset Purpose	The dataset will be used to train ML models to optimise the resource allocation along the network devices.
Dataset Type/Format	CSV file
Re-use of existing data	No
Dataset Origin	**Any** testbed
Dataset Collection	Deployment of monitoring probes over several UEs/ MEC nodes (mobile devices, terminals or tablets). Probes should measure the number of tasks under execution (per UE) and the available computation and storage resources available within a pre-defined period of time. Also, the total network traffic should be

	measured as the cumulative traffic generated by the end-devices.
Expected Size	Around 1-2 million samples
NANCY Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP4 and WP6
Relevant Tasks	T4.1, T6.5 – T6.9
Relevant Deliverables	D4.1, D6.5 – D6.9
Partners Services and Responsibilities	
Partner Owner	-
Partners responsible for data collecting	-
Partners responsible for data analysis	Bi2S
Partners responsible for data storage	Bi2S
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	YES
Naming Convention	Per UE/MEC node: Number of active tasks, computational resource utilization, storage resource utilization, network traffic (GBs)
Versioning	YES
Search Keywords	YES
Metadata	YES
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	YES
Methods/Tools for accessing the data	Download through a Zenodo repository
Access Restrictions	NONE
Repository	Not yet available
FAIR Data - Making interoperable	
Interoperability	Please describe if the Dataset will be interoperable.
Standards	Please provide any interoperability standards.
FAIR Data – Increase data re-use	
Licence	None required
Data sharing terms	Open data sharing
Dataset sharing after the end of the project	1 year after the project ends
Preservation and update of the dataset after the project	2 years after the project end
Re-use timeframe	Once within 6-month period
Allocation of Resources	
Cost for making the dataset FAIR	No cost is foreseen
Costs for the preservation (including backup) and the update of the dataset	For the backup: No costs are foreseen For updating the dataset: More experiments should be conducted in a NANCY testbed. Very small energy consumption costs are expected
Data Security	
Security Measures	None

Privacy Measure	None
Ethical Aspects	
Ethical and Legal Aspects	None
Consent form for data sharing and long preservation within the NANCY consortium	The consortium should sign a consent form in order to make the data open access.
Consent form for sharing the dataset publicly.	No consent form is required
Other Issues	No other issues are identified