

DIGITAL-ECCC-2022-CYBER-B-03-
CYBER-RESILIENCE



This project has received funding from the European Union's DIGITAL-ECCC-2022-CYBER-B-03-CYBER-RESILIENCE programme under grant agreement No 101145872.

Connected 5G-IoT Cyber Range for Training and Secure Operations



WP1– Project management

Deliverable D1.2 “Data managemenet plan - first version.”

Editor(s): UPRC

Author(s): Christos Xenakis (UPRC), Aggeliki Panou (UPRC), Aikaterini Poupouza (UPRC), Georgopoulos Nikolaos (UPRC), Styliani Tsitsoula (HCSI), Emmanouil Magkos (IURC), Ntantogian Christoforos (IURC), Anastasios Tzoumpas (UBI), Konstantinos Poulakakis (UBI), Ioannis Siachos(UBI), Roumelas George (UNILU)

Dissemination Level: PU

Type: DMP

Version: 1.3

Project Profile	
Contract Number	101145872
Acronym	NITRO
Title	Connected 5G-IoT Cyber Range for Training and Secure Operations
Project Starting Date	1 January, 2024
Project End Date	31 December 2026
Project Duration	36 Months

List of participants

1	 UNIVERSITY OF PIRAEUS	University of Piraeus Research Center	Greece
2		UNISYSTEMS LUXEMBOURG SARL	Luxembourg
2.1		UNI SYSTEMS SYSTIMATA PLIROFORIKIS MONOPROSOPI ANONYMI EMPORIKI ETAIRIA	Greece
3		IONIAN UNIVERSITY	Greece
4		HELLENIC CYBERSECURITY INSTITUTE	Greece
5		UBITECH LIMITED	Cyprus



Document History

VERSIONS

Table 1: Document history

Version	Date	Author	Remarks
0.1	01/05/2024	UPRC	Initial ToC and assignments
0.2	05/05/2024	UPRC	Updated sections 1 and 2
0.3	10/05/2024	UPRC	Updated section 3.1
0.4	15/05/2024	UPRC	Updated sections 4.1 and 5
0.5	21/05/2024	UPRC	Updated section 4.2
0.6	04/06/2024	IURC	Updated section 3.2.1 - NITRO data sources for Cyber Security Training Exercises
0.7	04/06/2024	HCSI	Updated section 3.2.3
0.8	10/06/2024	UBI	Updated section 3.2.2
1.0	14/06/2024	UPRC	First version
1.1	24/06/2024	UPRC	Review
1.2	27/06/2024	UPRC	Revised version
1.3	28/06/2024	UPRC	Final version

Table of Contents

Table of Figures	5
Table of Tables	6
Table of Abbreviations	7
Executive Summary	8
1 Introduction	9
2 NITRO Fair Data Principles	10
2.1 Data Summary	10
2.1.1 To be Findable	11
2.1.2 To be Accessible	13
2.1.3 To be Interoperable	14
2.1.4 To be Reusable	16
2.2 Allocation of Resources	16
2.3 Data Sharing	17
2.3.1 Methods of Data Sharing	18
2.4 Data Archiving and Preservation (including storage and backup)	19
2.5 Data Security	19
2.5.1 Data Integrity and Quality	20
2.5.2 Data storage	21
2.5.3 Data privacy	21
3 NITRO Data Sourcing and Data Sharing	22
3.1 Overview of Research Objectives (ROs') Scenarios	22
3.2 NITRO Data Sources	23
3.2.1 Cyber security training exercises	23
3.2.2 Scoring system and gamification	24
3.2.3 Use Cases	27
4 Ethical Aspects and Privacy and Security Requirements	28
4.1 Ethics	28
4.2 Legal framework	30
4.2.1 General Data Protection Regulation (GDPR)	31
4.2.2 Security and Authentication Legislation	31
4.2.3 EU AIA ACT	32
4.2.4 Cyber Resilience Act	33
5 The NITRO FAIR Dataset Template Questionnaire	34
References	43



Table of Figures

Figure 1: File share structure.	12
--------------------------------------	----



Table of Tables

Table 1: Document history.....	3
Table 2: Scoring system and gamification data.	26
Table 3: Needed Data for Use Case 1.	27
Table 4: Needed data for Use Case 2.....	28
Table 5: NITRO Project’s detailed information.	34
Table 6: NITRO’s FAIR Dataset Template Questionnaire.	36

Table of Abbreviations

Abbreviation	Explanation
DMP	Data Management Plan
AI	Artificial Intelligence
CRA	Cyber Resilience Act
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
PU	Public
SEN	Sensitive
ROs	Research Objectives
IoT	Internet of Things
5G	5 th Generation
6G	6th Generation
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
ERC	European Research Council
APIs	Application Programming Interfaces

Executive Summary

This deliverable presents the first version of the NITRO project's Data Management Plan (DMP). In more detail, the NITRO data management plan consolidates and structures the project's processes, procedures, and activities to support an effective and efficient data management methodology associated with the project schedule, needs, and scope. The deliverable outlines how the research Data collected/generated will be handled during and after the NITRO project, describes which standards and methodology for Data collection/generation will be followed, and whether and how Data will be shared. More specifically, the document will analyse the methods applied in line with FAIR data management to make data findable, openly accessible, interoperable, re-useable, and secure. It will also present detailed information on the project data lifecycle, privacy, and the project's policies for data collection, storage, access, sharing, protection, retention, and destruction. Furthermore, the legal framework, risks, and measures associated with ethical aspects, as well as mechanisms for data protection, governance, and trust, are addressed.

The document is based on the template and guidelines provided by the European Commission [ORD] [GFDM].

1 Introduction

The "D1.2 Data Management Plan" records the established protocols and standards that ensure privacy and integrity in all project activities and deliverables, in addition to the project's data management strategy. It outlines the NITRO strategy, the establishment of essential guidelines, and the framework for responsible and diligent data administration and research throughout the project. The activities and guidelines that arise from this Data Management Plan have been established in close collaboration with all consortium partners and will be rigorously adhered to across all work packages. Effective data management will enhance research results by increasing their sustainability and improving the validity and integrity of research methods.

NITRO endeavours to simultaneously accomplish the subsequent characteristics: user privacy, the flexibility of contract expression, security, and efficiency. This deliverable outlines a comprehensive set of guidelines addressing all essential subjects that require attention. The guidelines encompass a wide range of topics, including but not limited to the following: sharing strategies, infrastructures, applicable policies, roles, standards, data processing, storage, retention, and structure, adherence to legal requirements and market standards, best practices in ethics and privacy, identification, accessibility, and intelligibility, and legitimate utilisation for alternative objectives. These guidelines will be adopted during the initial phases of the undertaking. Furthermore, it establishes some procedural mechanisms for participants with the responsibilities of Data Controllers and Processors.

Through the NITRO project, Data Controllers and Processors are defined as below [GDPR]:

- **Data Controller** means the natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;
- **Data Processor** means a natural or legal person, public authority, agency, or other body which processes Personal Data on behalf of the controller;

The NITRO DMP outlines the fundamental components of the data management policy that the project participants shall implement for all datasets. This DMP also describes a strategy for managing research data during and after the completion of the project, including proposed methodologies and standards, data categories to be collected and processed, and so forth. Detailed information regarding the data lifecycle, privacy, and project policies related to the collection and processing (including storage, access, sharing, retention, and erasure) of data by NITRO can also be found.

The DMP adheres to internal ethics in order to ensure sufficient data protection measures and adherence to ethical standards. NITRO gathers data, which may contain personal information, from various sources. The data that we compile is categorised as follows:

- Qualitative and quantitative research data.
- Administrative data, such as information regarding participants, communications, and identity management.

- Information obtained from publicly available sources, such as legislation, government guidance, codes of practice, and ethical horizon scanning results.
- Open-source information gathered from sources accessible to the public.
- Publications and dissemination data, including information about proceedings, open peer-reviewed publications, interviews, reports, seminars and workshops, contact details for dissemination contacts, and inquiries.

The rest of this document is structured as follows.

- Section 2 describes how we plan to make the data FAIR internally and externally.
- Section 3 describes the initial data sets we have identified at the beginning of the project and the internal methods we use to store, exchange, and follow the data.
- Section 4 defines the Ethical Aspects and Privacy and Security Requirements.
- Section 5 concludes the report with the FAIR Dataset Template Questionnaire.

The present deliverable is of a public nature.

2 NITRO Fair Data Principles

2.1 Data Summary

For European Research Council projects, for each dataset, the following aspects should be considered:

- Making Data Findable,
- Making Data openly Accessible,
- Making Data Interoperable,
- Increase Data Re-use,
- Allocation of recourses and Data security

Specifically, as a component of ensuring that research data is discoverable, accessible, interoperable, and reusable (FAIR), this data management plan encompasses details regarding:

- The management of research data throughout and following the conclusion of the project;
- The specific data that will be gathered, processed, and produced;
- The methodologies and standards that will be employed;
- The decision of whether the data will be shared and/or made openly accessible;
- The procedures for curating and preserving the data.

The ownership of the Data collected/generated during the project will be held by the parties who have contributed to its production. The data owners will determine the availability level and the limits on its re-use on a case-by-case basis. The partners will adhere to the Findable, Accessible, Interoperable, Reusable standards. These guidelines require that data be made as accessible as feasible while ensuring that it does not harm the partners' commercial benefit. The Data will be disseminated among partners via internal repositories or direct communication with the public via the project's website or public repositories. The data will be retained for a minimum of 3 years and a maximum of 5 years after the project concludes. This will be done in accordance with the internal policies of each partner, who will store the data in their own repositories and cloud infrastructures.

The NITRO DMP should be regularly updated throughout the project to reflect significant changes, including but not limited to:

- Incorporating new datasets
- Modifying consortium policies (e.g., innovation potential, the decision to seek a patent)
- Adjusting consortium composition and responding to external factors (e.g., new members joining or existing members leaving).

The NITRO Databases will be stored in internationally recognised data repositories to ensure continuous availability if maintenance of the main web platform is terminated.

The project will employ Zenodo [ZND] as the repository for both storing and disseminating data. This repository will be linked to the NITRO project site on OpenAIRE. Zenodo is an innovative and intuitive platform that enables collaboration and presentation of research findings, including data and publications, for academics, scientists, EU projects, and institutions. It is particularly useful for sharing research not currently stored in institutional or subject-specific archives.

To adhere to the principles of making research data Findable, Accessible, Interoperable, and Reusable (FAIR), a DMP should satisfy the criteria outlined below [DMEC].

2.1.1 To be Findable

This section provides an overview of the standard procedures and practices used in the NITRO project for managing documents and data. The system consists of the NITRO file share that is integrated with Microsoft Teams. It also contains GitLab for software version control and internal templates with associated document metadata. For this concept to be implemented, the DMP must ensure that the following actions occur:

- Data and metadata are given a globally unique and permanently persistent identifier.
- Data is accompanied by extensive metadata.
- Data and metadata are recorded or catalogued in a resource that can be easily searched.
- Metadata provides a specific identity for the data.
- Describe the ease of finding data through metadata provision.
- Describe the capacity to identify data and reference standard identification mechanisms. Do you utilise persistent and distinct identifiers, such as Digital Object Identifiers (DOI)?
- Discuss the naming conventions employed.
- Establish the methodology for searching keywords.
- Provide guidelines for the production of Metadata, if applicable. If there are no established standards in your field, please explain the specific type of Metadata that will be generated and the process by which it will be developed.

2.1.1.1 Document Management Fileshare

The NITRO document management approach aims to reduce project participants' workload in creating, storing, locating, and synchronising documents. In this context, the Microsoft Teams solution is utilised for the purposes of document management, storage, and online creation and editing. The content contains all administrative details, document templates, project meeting planning and documentation, information regarding action points, and ongoing documents organised according to

work packages. Access to Microsoft NITRO Teams is provided through individualised email accounts. The graphic below illustrates the hierarchical layout of the main NITRO Teams folder.

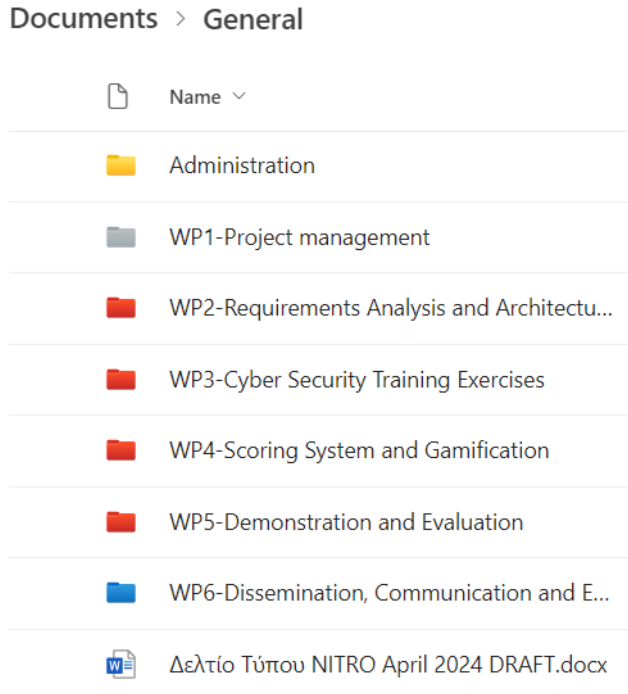


Figure 1: File share structure.

Provided below is a concise overview of the contents of each primary folder:

- **Administration:** Manage and organise essential project information, such as proposal documents, grant contracts, consortium agreements, project partner contact details, and the comprehensive project budget and associated activities.
- **Work Packages:** Each work package includes subfolders for the associated tasks, which contain all necessary content such as literature, work planning, and ongoing and finalised deliverables.

Within the NITRO project, the primary software utilised for creating documents is Microsoft Word, Excel, and PowerPoint, which are all components of the Microsoft Office suite. Microsoft Teams enables seamless collaboration on documents through its online editing feature, ensuring efficient joint usage.

2.1.1.2 Document metadata

Deliverable confidentiality information (dissemination level)

The NITRO project deliverables are categorised into two distinct distribution levels. *Public (PU)* deliverables are accessible to everyone. *Sensitive (SEN)* deliverables have limited access, subject to the conditions outlined in the NITRO Grant Agreement. The Grant Agreement has established the dissemination levels for all NITRO deliverables.

Deliverable status information

The following labels are used to indicate the state of the deliverable:

- *Internal draft:* A preliminary version of a deliverable that is still being worked on and is not yet ready for review;

- *Consortium approved*: A version of a deliverable that has been accepted by the internal reviewers of the project and is now ready to be submitted to the granting authority for review;
- *Final version*: A deliverable that has been accepted by the granting authority and is now ready for publication, taking into account the level at which the deliverable will be disseminated

Naming convention and versioning

The deliverable names within NITRO are intentionally crafted to possess both significance and distinctiveness, achieved through the utilisation of the following components:

- The project is named NITRO.
- The deliverable is identified per the NITRO contract (e.g., D1.2).
- The deliverable version begins with the deliverable version, which includes the table of contents (v0.1) and is labelled as the final version of the deliverable (v1.0). Subsequent iterations of a deliverable following a submission are denoted by an incremented version number (e.g., v1.1), with the following submission labelled v2.0, and so on.

A PDF created to submit a deliverable to EUSPA should have the same filename as the original document, except for the file extension.

2.1.1.3 Data management

Technical data

The technical data will be classified based on the confidentiality level and the data type. The amount of secrecy can be categorised as restricted to internal use only, limited to consortium usage only, and open to the public. The data management solutions utilised in the project directly secure those characteristics. The categorisation of data types is determined by factors such as the level of information they include, the level of development they have reached, and the specific format in which they are presented. Types such as source code, reference measurement data, processed data, analysis, and presentations can be distinguished, along with metadata. The ability to store and retrieve various types of data is contingent upon access to the storage platforms (requiring authentication), the use of licensed applications to read and write the data, and the availability of media exchange methods, such as the internet or memory storage (enabling access to the media).

Field test data

The project's beneficiary, Hellenic Cybersecurity Institute (HCSI), will manage the field test data by engaging members of the Hellenic Cyber Security Team, the Greek National Team that participates in the European Cyber Security Challenge (ECSC).

2.1.2 To be Accessible

To apply the principle of accessibility, the DMP must ensure that the following actions take place:

- Data are retrievable by their identifier using a standardised communications protocol.
- The protocol is open, free, and universally implementable.
- The protocol allows for an authentication and authorisation procedure, where necessary.
- Metadata is accessible, even when the Data are no longer available.

NITRO Database will be accessible through the NITRO project for three (3) to five (5) years following the project's end. During this period, unless otherwise decided by the consortium members, the Database functionality will remain the same as during the project duration.

2.1.2.1 Software versioning and revision control system

A separate GitLab instance is utilised to version the project's source code. The NITRO's GitLab repository is available here [NGT] (www.nitro-gitlab.ds.unipi.gr). The folder structure is initiated with each consortium member to allow for unrestricted code uploads. Code declaration will be carried out based on the available GitLab solutions, and file access protection will be implemented using restrictions and at a personal user level.

2.1.2.2 Dissemination of results

The NITRO team is well-informed about the open access policy that applies to scientific publications, as outlined in Article 29.2 of the Horizon Grant Agreement on Open Access to Scientific Publications. All peer-reviewed articles resulting from the research will be freely and openly accessible through an online repository and the project's website. The project will take the following actions:

- All project presentations, contributions, and publications will include the project logo and the required meta-data specified by the EC. This includes acknowledging the grant agreement number, call identifier, project name, and publication date.
- The project's publications will be uploaded to research-oriented social networks such as ResearchGate and open-access repositories like OpenAire and Zenodo no later than 6 months after their original publication date.
- Software not protected by intellectual property rights will be licensed as open source and openly distributed, for example, through platforms like GitHub.

2.1.3 To be Interoperable

In order to implement the notion of interoperability, the DMP must ensure that the following actions are carried out:

- Data/Metadata utilise a formal, easily understandable, collaborative, and widely applicable language for representing knowledge.
- Data/Metadata utilise vocabularies that adhere to the FAIR principles.
- Data/Metadata contain explicit references to other Data/Metadata.

At present, there is no broadly recognised protocol for facilitating the Data exchange between researchers, institutions, organisations, countries, etc. This includes adhering to established standards for Data annotation, Data exchange, compatibility with existing software applications, and enabling the merging of Datasets from various sources. Therefore, a human interpretation of the Data structure is always necessary in order to manually generate a Data map. Using standardised methods for data capture and well-documented annotation will facilitate the Data exchange.

More specifically, regarding interoperability, the following should be applied.

2.1.3.1 Overall interoperability

Interoperability encompasses the use of diverse methodologies and procedures to guarantee the seamless functioning of various systems and applications across technological, semantic, and organisational dimensions.

Related activities in the NITRO project are:

- Standardization of standards and protocols: Establishing a consistent and universally acknowledged set of standards and procedures would streamline the sharing and manipulation of data.
- Standardized data models: Developing standardised data models and information semantics will enable systems or applications to comprehend and interpret data accurately.
- Change management: Effectively managing changes and updates will ensure compatibility and consistency.
- Conducting interoperability testing enables the identification of potential issues and ensures seamless collaboration and coordinated efforts. General interoperability methods are essential for the project's cooperation to be effective and efficient.

2.1.3.2 Internal interoperability

Effective information processing and exchange in the NITRO project heavily relies on internal interoperability, which involves the integration of diverse technical and programming solutions. The following is a description of the steps to guarantee data compatibility in the project:

- Standardization of data formats: Establishing a uniform data format that will be universally adopted by all systems or apps involved in the project would streamline data sharing among them.
- Utilisation of open standards: Utilising open standards and protocols will guarantee system compatibility and simplify integration.
- Development of application programming interfaces: Implementing APIs will provide a seamless and protected data exchange across different systems.
- Consolidating data processing methods, such as authorisation and authentication processes, would guarantee the uniformity and safeguarding of data processing.
- Establishing a unified database: Establishing a centralised database will streamline data management and guarantee information consistency.
- Performing interoperability tests: Interoperability tests can identify issues and ensure that the systems incorporated in the project will be able to communicate data seamlessly.

Ensuring the compatibility of data in the project is essential for the efficient processing and sharing of information within the project, which ultimately leads to the project's success.

2.1.3.3 External interoperability

External interoperability refers to the NITRO solution's capacity to establish connections with various third-party systems utilised by (possible) users, partners, and other ecosystem participants for data exchange and collaboration in operational scenarios. To accomplish external interoperability, it is necessary to ensure the consistency of data formats and exchange protocols. Additionally, coordination of business and technical activities is required between different systems along the value chain while also establishing the relevant interfaces and APIs.

2.1.3.4 Message interoperability

Messaging interoperability refers to the capacity of the systems within the project to communicate and share messages and information using established communication standards and protocols. To achieve interoperability of communications, it is imperative to define the standards and protocols utilised in the project and ensure that all systems and applications are correctly set to process and interpret messages accurately.

2.1.3.5 Data Analytics

To establish a solid foundation for the NITRO project's data analysis, the following actions are carried out:

- Ensuring data quality, which means having properly organised, precise, comprehensive, and current data, increases the likelihood of conducting valid analysis. Establishing data quality benchmarks, executing data purification procedures, and performing routine data upkeep are essential for data integrity.
- Data cataloguing involves creating a centralised repository that provides an overview of available data resources. This includes information on the data's origin, schema, and metadata. Developing a data catalogue facilitates the process of locating, utilising, and examining pre-existing data resources.
- Standardising the form and structures of pertinent data and utilising widely established data formats greatly facilitates their analysis.
- Version control facilitates monitoring data modifications over time, simplifying the process of tracking changes and ensuring that all users utilise the latest data version.
- Consistently providing well-structured and standardised documentation, such as data dictionaries and data usage guidelines, facilitates comprehension of how to utilise the data and its contents.

2.1.4 To be Reusable

Lastly, to incorporate the concept of reusability, the DMP must ensure that there is implementation of the following actions:

- Data/Metadata have a plurality of accurate and relevant attributes.
- Data/Metadata are released with a clear and accessible Data usage license to permit the widest re-use possible.
- Data/Metadata are associated with their provenance.
- Data/Metadata meet domain-relevant community standards.
- Acknowledge whether the data produced and/or used in the project is re-useable by third parties. If the re-use of some Data is restricted, explain why and for what period a Data embargo is needed.
- Data quality assurance processes.
- The length of time for which the Data will remain re-usable.

NITRO Data will be licensed under a Creative Commons license to the extent that it may be subject to such licensing (likely the "CC BY"). Applicable Data will become available at the end of the project. Other scientists and interested parties can utilise the data. Certain parts of the data may have been accessible earlier due to their publication in scientific journals. There will be no embargo period. The project has allocated funds specifically to cover the costs related to the open-access publication. This is done to facilitate dissemination and re-use of the project's results.

2.2 Allocation of Resources

The costs for Data preparation to be FAIR are unknown at this stage but will be estimated in the future. Expenses may include any extra costs related to the release and documentation of the repository, if

appropriate. The project will bear the data preparation and management costs throughout its duration.

UPRC, in its role as the Project Coordinator for NITRO, will be responsible for creating the upcoming versions of the DMP (Data Management Plan) and archiving and publishing data in repositories. No extra funding is available for Data management activities for individuals who choose to participate in the experiment. Expenses associated with making research data freely accessible will be considered eligible for funding as part of the grant, regardless of whether the project participates in the pilot program, as long as the general eligibility requirements outlined in the Grant Agreement are met.

UPRC is the proprietor of the Microsoft Teams instance and is also in charge of operating the document management system. As a result, UPRC is responsible for guaranteeing the uninterrupted provision and high standard of service. Data management is the duty of data owners, who determine which data to disclose, to whom, for what reason, and under what circumstances. Data provisioning for research purposes is guaranteed by implementing the appropriate procedures and utilising open-access repositories. The usage of this data shall be restricted to the specific purposes of research, development, testing, and demonstration activities carried out within the project's scope in compliance with the principle of minimising data.

Suppose personal data needs to be processed. In that case, an explicit confirmation will be implemented to explicitly state that the recipient has a legal basis for processing the data and that appropriate technical and organisational measures are in place to protect the rights of the individuals whose data is being processed.

2.3 Data Sharing

The Data Controller will determine the specific details for sharing Data, including access procedures, potential embargo periods, technical mechanisms for dissemination, required software and tools for re-use, and whether access will be unrestricted or limited to specific groups. Furthermore, the Data Controller will determine the specific location of the repository where the Data will be stored if it already exists and has been recognised. This will include specifying the type of repository, such as an institutional or standard repository for the discipline.

Throughout the initiative, each prospective user seeking access will be directed to submit a "Request" to the Dataset controller within the NITRO consortium. Therefore, any potential user that needs to gain access would be guided to:

- Submit a "Request" to the Dataset controller from the NITRO consortium.
This request will contain:
 - Full name
 - Organization and department
 - Email address
 - Description of intended use
- After reviewing the request, if the Data controller approves it, the user will receive an email with a unique link to verify the email address.

- Then the user is asked to agree to and sign the following terms of access: [RESEARCHER_FULLNAME] (the "Researcher") has requested permission to use the Dataset. In exchange for such permission, Researcher hereby agrees to the following terms and conditions:
 - The researcher shall use the Database only for non-commercial research and educational purposes.
 - The Data Controller makes no representations or warranties regarding the Dataset, including but not limited to warranties of non-infringement or fitness for a particular purpose.
 - The researcher accepts full responsibility for his or her use of the Dataset and shall defend and indemnify the Data Controller, including their employees, trustees, officers, and agents, against any and all claims arising from the Researcher's use of the Dataset, including but not limited to Researcher's use of any copies of copyrighted Dataset that he or she may create from the Dataset.
 - The researcher may provide research associates and colleagues access to the Dataset provided they first agree to be bound by these terms and conditions.
 - The Data Controller reserves the right to terminate the Researcher's access to the Database at any time and without justification.
 - If a for-profit, commercial entity employs the Researcher, the Researcher's employer shall also be bound by these terms and conditions, and the Researcher hereby represents that he or she is fully authorised to enter into this agreement on behalf of such employer.
 - The law and jurisdiction of the Data Controller's country shall apply to all disputes under this agreement.

2.3.1 Methods of Data Sharing

The choice of techniques for sharing data will rely on various aspects, including the data's nature, amount, complexity, and sensitivity. Data can be communicated using any of the following means [DSG]:

- **Under the auspices of the Principal Investigator**

Investigators who are authorised by the Principal Investigator may securely send Data to a requestor or post the Data on their institutional website. Investigators should contemplate utilising a Data-sharing agreement to enforce suitable restrictions on the secondary utilisation of the Data.
- **Through a third party**

Investigators have the option to share their Data by transferring it to a Data archive facility, which will then distribute it to the scientific community. This ensures wider access, preserves proper documentation, and fulfils reporting obligations. Data archives are especially appealing to researchers interested in efficiently handling a substantial number of data requests, evaluating trivial or unsuitable queries, or offering technical support to users looking to contribute to data studies.
- **Using a Data enclave**

Data sets restricted from public dissemination due to confidentiality concerns or third-party license or use agreements that forbid redistribution can be accessible via a Data enclave. A

data enclave is a highly protected and controlled area where authorised researchers can analyse using restricted data resources.

- **Through a combination of methods**

Investigators have the option to release their data using a mix of the mentioned ways or in other versions. This allows them to regulate the level of access granted.

Note: During the NITRO project, Data controllers could use a combination of methods for Data sharing.

2.4 Data Archiving and Preservation (including storage and backup)

Depending on the specific dataset, the respective data controller will be responsible for implementing data archiving and preservation protocols for long-term data preservation. This encompasses specifying the duration for which the Data should be retained, estimating its final volume, determining the associated expenses, and outlining the intended funding methods.

2.5 Data Security

Emerging technologies present numerous opportunities to gather, utilise, and exchange data more effectively. These opportunities include facilitating research and enhancing quality, safety, and efficiency. However, they present novel obstacles regarding safeguarding privacy and ensuring data security.

Regarding the Guide on Good Data Protection Practice [GGD], to process data in a secure manner, each Data Controller must:

- Implement both technical and organisational measures to prevent unauthorised access and
- Establish explicit access restrictions.
- Arrange the processes in a manner that provides optimal control, for example, by enabling access tracking through a logbook.
- If someone handles the data on your behalf, ensure that this processor implements adequate security measures.

Practically speaking, these measures could lead to:

- User authentication: The process of confirming the identity of a user.
- Access control: A system that determines whether a user is allowed or denied access to specific data.
- Storage security: Safeguarding data from unauthorised access by implementing measures such as operating system controls for authentication and access control. This could be achieved by:
 - Operating system controls (authentication & access control)
 - Use of passwords to access electronic files (e.g. use the text editor function to save a document password-protected)

- Local encrypted storage (enable the full disk encryption, enable the file system, enable the text editor encryption)
 - Database encryption: turning data into a form that makes them unintelligible (for anyone not having access to the key)
- Communication security: Secure electronic communication methods for exchanging data can include the following options:
 - Encrypted communication (SSL/TLS) (e.g., use web services whose URL starts with 'https: //' and not only http://);
 - Firewall systems and access control lists (e.g. make sure the firewall service is enabled on your PC);
 - Anti-virus & anti-malware systems;
 - Protect Data and Data carriers when they are physically transferred (paper notes, laptop, etc.).
- Additional IT technological controls include implementing security upgrades, anti-virus protection, local data backups, and the prevention of specific program installations.

2.5.1 Data Integrity and Quality

Ensuring data integrity and quality are crucial for GDPR compliance, and NITRO will employ many technical measures to achieve this, including:

- Data encryption: The method of converting information into a code that can only be decoded by the intended recipient. It safeguards sensitive data from illegal intrusion. The General Data Protection Regulation recommends using encryption to store or transmit sensitive personal information.
- Access control: Only authorised individuals are permitted to access sensitive information. The General Data Protection Regulation (GDPR) requires organisations to create access policies and procedures that limit access to data based on the principle of least privilege.
- Data backups: Regularly doing backups ensures the safety of personal information in case of a data breach or system failure. In order to adhere to the General Data Protection Regulation (GDPR), organisations must possess a mechanism for safeguarding and restoring confidential data.
- Data pseudonymisation: Data pseudonymisation refers to the process of replacing personal information with a code or alias. It enables data utilisation for research and analysis while reducing the probability of unauthorised access and data breaches.
- Data auditing and monitoring: Auditing and monitoring involves frequently tracking metrics related to data quality, including completeness, correctness, and consistency. It enables

organisations to promptly identify data quality issues and take corrective actions to address them before they jeopardise data security.

- Data validation and verification guarantee data accuracy, coherence, and validity. They involve comparing data with established norms and standards to ensure its quality and integrity.

2.5.2 Data storage

Data storage and possible smart contract implementations will use privacy-enhancing techniques to protect information secrecy. User data will be protected via privacy-preserving smart contracts, and a robust encryption method will be evaluated for data storage.

Practically, these procedures could lead to:

- Implementation of operating system controls, specifically authentication and access control.
- Passwords employment as a means of gaining access to electronic files (e.g., using the text editor function to save a document with password protection).
- Encryption for local storage, including complete disk, file system, and text editor.
- Database encryption involves transforming data into an unreadable form that can only be accessed by those who possess the cryptographic key.

2.5.3 Data privacy

A conceptual taxonomy for security and privacy will be utilised, encompassing the following principles of security and privacy:

- Data confidentiality: ensuring the privacy and protection of Personal Data.
- Data provenance: ensuring the integrity and validation of Personal Data.
- Ensuring the security and privacy of Big Data and meeting the necessary data protection regulations in public policy, societal, and cross-organizational contexts.

In order to ensure security, access control will be implemented so that only authorised clients can access and retrieve information from the ledger.

Every NITRO software module will have cutting-edge security, authentication, and authorisation techniques. The durability of these modules is guaranteed by extensive advancements in the field (the fundamental components originate from previously funded EU projects or existing commercial solutions) and will undergo rigorous penetration/hacking tests and challenges for validation. Furthermore, a collection of secure APIs and Smart Contracts will be provided to ensure the availability of Data protection solutions.

NITRO's internal monitoring function will combine scanning and tracking capabilities to detect stolen or illegally obtained data. This will be achieved by using fingerprinting or watermarking as a reactive way to uncover data leakage.

NITRO Data processing must be lawful, fair, and transparent. It should involve only necessary Data and be proportionate to achieve the specific task or purpose for which they were collected. Therefore,

NITRO will only collect the Data needed for the research objectives since collecting unnecessary/unrelated Data for the research project may be deemed unethical and unlawful. The Data are to be processed only for scientific purposes comprising processing operations performed for study and systematic research to develop scientific knowledge for the sector addressed by NITRO.

3 NITRO Data Sourcing and Data Sharing

3.1 Overview of Research Objectives (ROs') Scenarios

The interactions mapped in the Research Objectives (ROs) scenarios have determined the Data sources, as well as the connections that will take place in NITRO. A short description of each RO's scenario can be found below:

R01. Deliver a novel cybersecurity platform for virtualised 5G-IoT cyber range services.

One of the key contributions of the NITRO project will be to improve the security of standalone and interconnected 5G and IoT environments. To this end, the integration of existing testbeds for IoT and 5G will be considered in order to create a unique new environment for security training. In the upcoming development, at least four distinct cyber ranges and testbeds are slated to be interconnected. Within the NITRO cyber range framework, two specific sectors will be addressed: IoT and 5G operators. Moreover, a novel cyber range tailored explicitly for the telecommunications and IoT domains is also on the horizon.

R02. Create the first common Data Repository for 5G-IoT cyber ranges that will include attacks, threat analysis, scenarios for both 5G and IoT networks.

Open standards, such as XML and JSON, will be employed for interoperability to define the metadata structure. The NITRO knowledge repository will be a pioneering instrument for knowledge transfer and big data analytics playground, facilitated by a suitable open-data access policy. A minimum of one cyber range operator is expected to repurpose data from the repository, with the establishment of at least one common data repository. Furthermore, three cyber range operators are projected to repurpose data from this repository. Additionally, integration with at least two European initiatives and frameworks for cybersecurity information sharing and intelligence is planned.

R03. Offer a powerful training environment for improving the preparedness of professionals (from staff in security operation centres to pen testers) through a variety of exercises.

NITRO's innovative virtual cyber warfare training environment is designed to empower the next generation of cyber defenders, enabling them to enhance their cybersecurity skills and better prepare for the detection, blocking, and mitigation of emerging cyberattacks. This will be achieved through testing and evaluating security technologies and methodologies under high-stress, real-world conditions, including virtualised versions of IDS/IPS and penetration testing tools.

The initiative aims to develop at least one large-scale, cross-sector scenario between the IoT and 5G domains. Additionally, it plans to support at least 50 structured training and cybersecurity exercises for public and private organisations. Furthermore, NITRO intends to create at least two large-scale,

cross-sector scenarios between the IoT and 5G domains, supporting a minimum of 100 structured training and cybersecurity exercises for public or private organisations.

R04. Emulate innovative AI adversarial attacks that can drive impactful attacks on 5G and IoT network management structures along with detection capabilities testing.

This objective will involve emulating six AI adversarial attacks, three targeting the 5G domain and three targeting IoT. Additionally, twelve AI adversarial attacks will be simulated, six targeting the 5G domain and six targeting IoT.

R05. Expand the access to cyber-range facilities to a broader range of stakeholders.

This objective prioritises cyber defenders from all European regions, especially those lacking access to unified cyber range facilities. NITRO will extend its benefits to external users, targeting a minimum of 1000 cyber defenders initially granted access to its services, aiming to reach at least 2000 cyber defenders.

In addition to the above-described objectives, the NITRO initiative aims to enhance existing resources and introduce new functionalities to augment their capabilities. This includes achieving a minimum of a 25% capacity increase in the activities of two 5G and IoT cyber ranges, as well as a target of at least a 50% capacity increase in the activities of two 5G and IoT cyber ranges.

NITRO's functionality is going to be demonstrated through the organisation of two events. These two events are going to have the additional purpose of generating interest within the European cybersecurity community.

3.2 NITRO Data Sources

Thus, the NITRO's data sources issue is directly affected by the pillars of the NITRO infrastructure, which are the Cyber security training exercises and the Scoring system and gamification, as well as its use cases.

3.2.1 Cyber security training exercises

The cyber range scenarios for 5G cybersecurity training involve various types of data. The exercises will not require any personal data, however during the exercises, the following might be considered:

- **Personal Details:** Names, email addresses, and access credentials used within the training environment. Pseudonyms will be generated for the participants and the accounts will not keep any PII information except the email addresses. Passwords will be encrypted using strong encryption algorithms like Argon2, bcrypt, scrypt.
- **Network Data:** IP addresses, MAC addresses, timestamps, and network traffic data. It is required to emulate MAC addresses in that case and avoid using the real MAC address.
- **System Logs:** Detailed logs of system operations, security events, and user actions for auditing and analysis. Since MAC addresses will not be real, the system logs will not include privacy risks.

- Application Data: Information will be stored and processed by internal applications of the cyber range. The participants are not obligated to put any personal information.

The cyber range scenarios, include network data such as IP addresses, MAC addresses, timestamps, and traffic data; user data like user identification information, browsing history, and location data; and logs detailing system activity, security events, and user actions. These data are generated and used to simulate real-world cyber threats and train personnel on specific security scenarios.

The participants as data subjects are not obligated to share personal data, and after the participation, no collected data will be kept. A Data Protection Impact Assessment (DPIA) would be helpful to ensure the requirements for data processing in alignment with the privacy needs and to identify and mitigate potential privacy risks.

In case the collected logs and datasets are kept, privacy preservation algorithms will be used, removing personal identifiers from data before analysis to protect individuals' privacy. Data encryption, for example users' passwords and credentials, ensures that all personal data is encrypted to prevent unauthorized access and minimize privacy risks. In case of data collection for retrieving statistics from the participation or for research usage, informed consent will ensure that trainees are fully informed about how their data will be used and that they consent to this use.

A DPO is not necessary since there is no Personally Identifiable Information (PII) collection. This is because the cyber range scenarios are simulated environments used for cybersecurity training and testing. The data is not used for any real-world purposes, and the subjects are informed that no personal information will not be used in the training sessions.

3.2.2 Scoring system and gamification

Concerning the datasets to be created during the project activities towards the establishment of an "*attack playbook*", with various attack vectors, based on which the NITRO scoring system will be created, the following have been identified:

- Detailed analysis of the most prominent types of attacks that can be considered in the overall 5G/6G continuum. This will include identification of adversaries' capabilities trying to exploit all possible layers of the networking stack – including both the control and data planes but also considering host-based exploits targeting the virtual function chains deployed on network nodes (e.g., VNFs). This will set the scene for the compilation of an "*attack playbook*" comprising a break-down of threats and vulnerabilities that will be simulated as part of the cyber-range activities.
- Establishment of a set of cyber-range activities to be performed for both validating the security posture and resilience of 5G/6G environments but also as training exercises towards capability and expertise development. This will also be associated with access control mechanisms for enabling users to get (authenticated) access to the results of their activities.
- New algorithms and mechanisms will be developed (and partially implemented as proof-of-concepts) for the scoring of this set of cyber range activities. This will built on existing scoring

mechanisms fusing both the impact of an exploit performed with feasibility rating (classifying the difficulty level of the activity based on the level of complexity) but also based on the attack path that was exploited (e.g., single-target exploits vs. attack path exploitation spanning the entire service-graph-chain).

- Identification of security controls and mechanisms capable of protecting against launched attacks and exploits.
- New tracing mechanisms that will be used for extracting the system measurements (of the target nodes) as part of the verification/monitoring of the successful launch of an attacks as part of the cyber range activity.
- Datasets generated from the implementation and experimentation of these cyber range activities in the project evaluation testbeds.
- Metadata focusing on the implementation of the cyber-range activities. These will follow the structure of README file providing details on how to setup and deploy all environments created.

All of the above will be wrapped within a framework to be offered enabling the following functionalities:

- **Data Feeding the System:** The data fed into the scoring system and gamification system primarily includes:
 - User Interaction Data: This includes data on user actions, progress, achievements, and participation in various activities within the system.
 - Performance Metrics: Data on user performance in tasks, challenges, or exercises that are part of the gamified system.
 - Feedback Data: User feedback on tasks and activities, including ratings, comments, and suggestions.
 - System Logs: Logs that capture user activities, system responses, and any errors or anomalies.
- **Data Processing:**
 - Data Collection: Data is collected through user interactions with the system, including completion of tasks, participation in challenges, and feedback submissions.
 - Data Aggregation: Collected data is aggregated to provide a comprehensive view of user engagement and performance.
 - Analysis and Scoring: The data is analysed using predefined algorithms to generate scores, badges, and leaderboards. This may involve statistical analysis, machine learning models, or rule-based scoring systems.
 - Feedback Loop: The system uses the analysis results to provide real-time feedback to users, adjusting difficulty levels or suggesting new challenges based on performance.

Access to and sharing of any research data will be kept open-source for helping to advance science and to maximize the research investment. Data producers will take efforts to provide effective informed consent statements to respondents, to identify data before deposit when necessary, and to

communicate any additional concerns about confidentiality. With respect to timeliness of data deposit, archival experience has demonstrated that the durability of the data increases and the cost of processing and preservation decreases when data deposits are timely. It is important that data is deposited while the producers are still familiar with the dataset and able to fully transfer their knowledge to the archive.

Table 2: Scoring system and gamification data.

Type of data	Format	Volume	Management Tool
Service metrics (data generated during the implementation of the cyber-range activities)	Results on the scoring system that will be presented to the users as feedback on their progress during the implementation of the exercises Txt format	GB	NITRO Framework User Interface
Simulation data sets capturing the operation and result of the selected attack vectors to be simulated	JSON	MB	Gitlab
Software prototypes and exploit launching software abstractions for enabling NITRO's cyber-range activities	C-based libraries	MB	Gitlab
Results from experiments (in the context of the use cases) as well as the integration of the overall NITRO framework	Software packages		Gitlab
Simulated datasets for emulating difficult-to-deploy attack vectors	JSON format emulating the behavior of compromises virtual functions deployed	GB	Can be supported through Dockers
Documentation	.pdf, .xls, .ppt, jpg, png	MB	Microsoft Office
Threat Modelling and Risk Assessment methodology merging CVSS 0.3 and	Software Libraries	MB	Gitlab

Type of data	Format	Volume	Management Tool
attack potential rating as defined in ISO/SAE			

Need for a Data Protection Officer (DPO):

- Assessment: Given that the scoring system and gamification platform will process user interaction data, including performance and feedback, there might be a need to ensure compliance with data protection regulations such as GDPR.
- Recommendation: Appointment of a DPO may be necessary to oversee the processing activities, ensure compliance with data protection laws, and handle any data-related concerns or breaches.

Need for Data Protection Impact Assessment (DPIA):

- Assessment: Considering that the system involves the collection, processing, and analysis of user data, a DPIA is recommended to assess the potential risks to user privacy and data security.
- DPIA Process:
 - ✓ Identify Data Processing Activities: List all data processing activities involved in the gamification and scoring system.
 - ✓ Assess Risks: Evaluate the potential privacy risks and impacts on individuals' rights and freedoms.
 - ✓ Mitigation Measures: Propose measures to mitigate identified risks, such as data anonymisation, access controls, and encryption.
 - ✓ Documentation: Document the DPIA findings and the steps taken to address risks.

3.2.3 Use Cases

Lastly, the NITRO's data sources regarding its two distinguished use cases are presented in a coherent manner below.

Table 3: Needed Data for Use Case 1.

Use case 1: Cybersecurity Testing of 5G and IoT Applications and Network Services	
Pilot use case Lead Partner:	Ionian University (IURC)
Human Participation Details	
Recruitment	Students; IT Professionals; Cybersecurity Professionals; Research Workforce; IoT device manufacturers; Network Administrators
Benefit of the participants	Hands-on Experience; Cybersecurity Testing and Training; Enhance Skills and Knowledge of 5G and IoT
Recruitment Place:	Open calls; Workshops; Conferences; Online Platforms; Social Media, and Professional Networks
Recruitment Responsible:	Depending on the exercise

Exclusion Criteria:	Not defined
Participant Type:	Students; IT Professionals; Cybersecurity Professionals; Research Workforce; IoT device manufacturers; Network Administrators
Data Type:	IP Addresses; MAC Addresses; Timestamps; Traffic Data; IoT device data; Usernames; Passwords; System Logs; Application data
Participation Type:	Active participation in cybersecurity testing and training; Passive participation in demonstration and seminar
Participation Place:	Online Workshops; Workshops in Person
Risks:	No potential privacy risks

Table 4: Needed data for Use Case 2.

Use case 2: Security Training for ECSC	
Pilot use case Lead Partner:	ELLINIKO INSTITOUTO KYVERNOASFALEIAS ASTIKI MI KERDOSKOPIKI ETAIREIA (Hellenic Cybersecurity Institute - HCSI)
Human Participation Details	
Recruitment	Members of the Hellenic Cyber Security Team;
Benefit of the participants	Practical Experience; Cybersecurity Testing and Training; Improve Expertise and Understanding of 5G and IoT
Recruitment Place:	ECSC participants
Recruitment Responsible:	Depending on the objective
Exclusion Criteria:	Not defined
Participant Type:	Students between 14 and 25 years old participating in a national cybersecurity team, taking place in the ECSC
Data Type:	IP Addresses; MAC Addresses; Timestamps; Traffic Data; IoT device data; Usernames; Passwords; System Logs; Application data
Participation Type:	Active participation in the technical validation of the NITRO Cyber Range in real cases and its performance monitoring, as well as the obtention of baseline measurements of KPIs. Passive participation in the improvement of the technical aspects and usability aspects of the platform through questionnaires specifically created for this purpose.
Participation Place:	Online workshops; Workshops in Person
Risks:	No potential privacy risks

4 Ethical Aspects and Privacy and Security Requirements

4.1 Ethics

Data protection and privacy are fundamental rights that must be safeguarded. It relates to the technical and legal structure designed to protect Personal Data against unanticipated, malicious, or accidental utilisation. As a result, data protection encompasses protocols that concern data collection, access, transmission, and preservation, among others. Furthermore, a data protection strategy may encompass protocols to ensure the data's precision. Privacy concerns emerge in the realm of research whenever personally identifiable information is gathered and retained, whether in digital or physical format. The main challenge for research is to use and share the data and, at the same time, protect personal privacy [GGD].

Regarding EC Ethics and Data Protection, in a research and development context, measures to achieve data protection by design could include:

- Data Confidentiality: NITRO partners must keep any data, documents, or other material confidential during the project's implementation and for five (5) years after the project's completion.
- The pseudonymisation or anonymisation of Personal Data:

When data administrators must maintain a connection between research subjects and their Personal Data, it is recommended that the Data be pseudonymised whenever feasible. This measure aids in protecting data subjects' privacy and mitigates the risk of infringement upon their fundamental rights in the event of unauthorised entry. Pseudonymisation preserves the identity of the data subject by combining supplementary identifiers with the pseudonym (e.g., code number or key code). The time and effort required to identify the individual, as well as the available technologies, are decisive for determining whether it is possible to identify the data subject from pseudonymised data.

On the other hand, a practical approach to tackle the ethical concerns surrounding personal data involves the process of anonymisation, which entails removing any identifying information [EDP]. Non-personal data, including statistical and collective information and data anonymised to the extent that it is impossible to identify the source, do not qualify as personal data and are thus exempt from the protection of data protection regulations. However, significant ethical concerns may still emerge even when utilising anonymised datasets exclusively, rendering the database highly impractical. The potential for re-identification (the process of transforming pseudonymised or anonymised data back into personal data through data matching or similar techniques) makes anonymisation an increasingly difficult task. Merely eliminating direct identifiers does not guarantee that the data subject can no longer be identified, and additional measures are usually necessary to prevent identification, depending on the context and purposes of the processing for which the anonymised data are intended. For these purposes, NITRO will apply standard Anonymisation techniques such as [GGD]:

- Randomisation: A family of techniques removing the link between the Data and the individual. If the Data are sufficiently uncertain, then they can no longer relate to a specific individual.
- Generalisation: An approach consisting of generalising or diluting the attributes of Data subjects by modifying the respective scale or order of magnitude (i.e. a region rather than a city, a month rather than a week). Generalisation can effectively prevent singling out, but it does not allow adequate anonymisation in all cases.
- Data minimization: Data processing must be lawful, fair, and transparent. It should involve only Data that are necessary and proportionate to achieve the specific task or purpose for which they were collected. Therefore, NITRO will only collect the Data needed for the research objectives since collecting unnecessary/unrelated Data for the research project may be deemed unethical and unlawful.
- Applied cryptography (e.g. encryption and hashing);
- Using Data-protection focused service providers and storage platforms;

4.2 Legal framework

In addition to the previously mentioned general conditions, the consortium acknowledges that several privacy and data protection concerns may be caused by the activities to be performed in the project's scope. Consequently, in the event that specific parts of the project require the participation of human beings, such involvement shall be conducted in strict adherence to all applicable European and national legislation and directives related to the jurisdiction in which the data collection is occurring.

The NITRO partners found the following regulations and policy documents to be relevant and considered when dealing with personal data:

- The Universal Declaration of Human Rights [DHR] and the Convention 108 [COE] for the Protection of Individuals about Automatic Processing of Personal Data;
- Directive 95/46/EC [LUS] & Directive 2002/58/EC [EPD] of the European Parliament regarding issues with privacy and protection of personal data and the free movement of such data;
- Charter of Fundamental Rights of the European Union;
- European Convention for the Protection of Human Rights and Fundamental Freedoms;
- Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data;
- Directive 2002/58/EC on data protection.

Specifically, when handling personal data, the NITRO partners will adhere to the following guidelines:

- Personal data is only gathered with specific consent from the individuals involved. The personal data collected is treated confidentially and carefully, using appropriate technical measures for information protection;
- The personal data needed for user requirements and market analysis, as well as any statistical assessment is processed anonymously, i.e., without association with the names of individuals;
- The gathering of data related to individuals exclusively based on their racial heritage, religious beliefs, sexual orientation, political stance, or membership in specific movements or organisations that are not legally prohibited is strictly prohibited and is not conducted as part of the project;
- The research ensures that any shadow (ancillary) personal data acquired is promptly deleted. The ultimate goal, nevertheless, is to reduce the amount of such supplementary information to a minimum. Compliance with Art.2 of the Committee of Ministers Recommendation R(87)15 of the Council of Europe regarding the regulation of the use of personal data in the police sector is also given special consideration;
- No data collected is sold or used for purposes other than the current project.

Data administrators must disclose pertinent details regarding the initial data collection process, methodology, and informed consent procedure if consent is the suitable legal foundation when utilising Personal Data from a prior research endeavour. Additionally, they must ensure that they comply with the data protection principles and have obtained the data controller's consent before using the data in the NITRO project.

Transferring personal data outside the European Union can increase the risk of unauthorised access, disclosure, or use of this information, making it harder for individuals to protect their data protection rights. EU Member States' national authorities must cooperate and exchange personal data to perform their duties or assist authorities in other Member States. Effective data protection is essential to uphold the EU's values and principles, requiring international cooperation and agreements. The European Data Protection Supervisor (EDPS) regularly communicates with Data Protection Authorities (DPAs) and regulators within the EU and globally to enhance cross-border enforcement.

4.2.1 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR), implemented on May 25, 2018, is an all-encompassing collection of data protection regulations from the European Union. The GDPR applies to all entities, including individuals, enterprises, and organisations, irrespective of their location, that process the personal data of EU residents.

Its objective is to safeguard individuals' fundamental rights and liberties, which encompasses preserving their personal data and the right to privacy. Before collecting individuals' data, the regulation mandates that organisations secure their informed consent and guarantee that the data is processed legally, equitably, and transparently. A person has the right, per the GDPR, to be informed of the nature, purpose, and intended use of any private information being collected about them. In addition, they may object to processing their data and have the right to access, rectify, and eliminate it.

The GDPR mandates data protection officers, data protection impact assessments, and data incident notifications, imposing stringent obligations on organisations that process personal data. Furthermore, organisations must implement appropriate organisational and technical safeguards to prevent unauthorised access, disclosure, or deletion of personal information.

Additionally, the regulation profoundly affects scholars who gather, manipulate, and distribute personally identifiable information. Researchers are obligated to obtain participants' informed consent, provide a comprehensive explanation regarding the intended use of their data, and guarantee that the data is processed in adherence to the regulations outlined in the GDPR. Additionally, researchers are obligated to verify the compliance of any third-party vendors or partners who handle personal data on their behalf with the GDPR.

GDPR is a comprehensive set of data protection regulations that imposes substantial obligations on project team members who process personal data in the context of NITRO. When collecting, processing, and disclosing personal information, they must ensure compliance with GDPR requirements to avoid legal repercussions and reputational harm.

4.2.2 Security and Authentication Legislation

- The Directive (EU) 2016/1148 on Network and Information Security (NIS Directive) provides legal measures to boost the overall level of cybersecurity in the EU and is the first piece of EU-wide cybersecurity legislation. The NIS Directive aims to establish a minimum level of (cyber) security for network and information systems across the EU, particularly for those operating essential services. The Directive explicitly addresses operators of basic services and digital

service providers. However, it is up to the Member States to assess which entities meet the criteria of the definition of an operator of an essential service. Member States must identify the operators of essential services.

- The European Parliament adopted the Regulation on ENISA, the "EU Cybersecurity Agency," repealing Regulation (EU) 526/2013 and on Information and Communication Technology cybersecurity certification (Cybersecurity Act) [REP] on March 12, 2019. This Act aims to strengthen Europe's cybersecurity by replacing existing national cybersecurity certification schemes with European schemes that will define security objectives. For one thing, NITRO will comply with the Cybersecurity Act's security principles by design and by default.

4.2.3 EU AIA ACT

The Artificial Intelligence Act (AI Act), proposed on April 21, 2021, is a comprehensive set of regulations introduced by the European Union to govern the development and deployment of artificial intelligence technologies. The AI Act applies to all entities, including individuals, enterprises, and organisations, irrespective of their location, that provide or use AI systems within the European Union.

Its objective is to ensure the safe and ethical development and use of AI, safeguarding fundamental rights and freedoms, including privacy, non-discrimination, and protection from harm. The regulation mandates that AI systems be developed and deployed in a transparent, traceable, and accountable manner. Before deploying AI systems, organisations must conduct risk assessments to identify and mitigate potential harms, ensuring that these systems comply with legal, ethical, and technical standards.

Under the AI Act, organisations are required to inform users about the nature, purpose, and functioning of the AI systems they interact with. Users have the right to receive clear and understandable information about the AI system's decision-making processes and the logic behind them. Additionally, individuals can object to decisions made by AI systems and request human oversight, especially in high-risk scenarios. Furthermore, researchers are obligated to obtain informed consent from participants, provide detailed explanations regarding the use and impact of AI systems, and ensure that their research adheres to the principles outlined in the AI Act. Additionally, researchers must verify the compliance of third-party partners and vendors involved in their projects with the AI Act.

The AI Act represents a comprehensive framework for regulating AI, imposing significant obligations on project teams developing or deploying AI technologies in the EU. Compliance with the AI Act is essential to avoid legal penalties and to foster trust and accountability in AI applications.

As of March 2024, significant progress has been made towards implementing the AI Act. The European Parliament and the Council of the European Union reached a provisional agreement on the final text of the AI Act, paving the way for its formal adoption. This agreement includes refinements to the risk-based classification system, clearer guidelines on the requirements for high-risk AI systems, and enhanced measures to protect fundamental rights. Additionally, the EU has initiated a series of pilot projects and regulatory sandboxes to help organisations and researchers test and refine their AI systems in compliance with the new regulations. These initiatives aim to facilitate a smoother

transition and provide practical insights into applying the AI Act's provisions.

Based on the comprehensive analysis of the NITRO project's goals and tasks, it is evident that the project should comply with the AI Act. A significant aspect of NITRO is delivering adversarial AI exercises aimed at researching and utilising relevant attack methods that exploit vulnerabilities in AI models within 5G and IoT networks.

Given the high-risk nature of these activities, especially in critical infrastructure areas like 5G and IoT networks, the NITRO project should comply with the AI Act. This involves several key measures:

- **Risk Management:** Establish a detailed risk management framework, including continuous monitoring, assessment, and mitigation of risks associated with adversarial AI in 5G and IoT networks.
- **Transparency and Accountability:** Maintain comprehensive documentation of adversarial scenarios, testing methodologies, and defensive strategies to ensure transparency and explainability of AI processes.
- **Robustness and Security:** Implement rigorous testing protocols to evaluate and enhance the robustness and security of AI models, ensuring they can withstand adversarial attacks.
- **Human Oversight and Control:** Incorporate mechanisms for human oversight to monitor AI system performance, intervene when necessary, and ensure AI systems behave as intended under adversarial conditions.
- **Data Governance and Privacy:** Adhere to strict data governance policies to ensure the quality, integrity, and privacy of data used in training and testing AI models, complying with GDPR and other relevant data protection regulations.
- **Audit and Reporting:** Conduct regular compliance audits and prepare detailed reports to demonstrate adherence to the AI Act, ensure ongoing compliance, and address emerging regulatory requirements.

4.2.4 Cyber Resilience Act

The Cyber Resilience Act (CRA) is a legislative framework introduced by the European Union to enhance the cybersecurity and resilience of digital products and services within the EU market. The primary objectives of the CRA are to ensure cybersecurity, enhance resilience, and promote trust in digital products and services. It establishes baseline security requirements to protect against cyber threats, improves the ability of critical infrastructure and digital services to withstand and recover from cyber incidents, and increases consumer and business confidence in the security and reliability of these products.

The CRA applies to a broad range of digital products and services, including hardware and software that are connected to the internet, such as IoT devices and applications, critical infrastructure like 5G networks, and supply chain components integrated into critical systems. To comply with the CRA, organisations must meet several key requirements. These include implementing comprehensive risk

management processes, integrating security by design into product development, having effective incident response and recovery plans, conducting regular vulnerability assessments and penetration tests, ensuring supply chain security, maintaining transparency and accountability, and performing regular compliance audits.

Given the scope and objectives of the CRA, the NITRO project should adhere to these requirements to ensure compliance. The NITRO project, focused on creating a 5G-IoT cyber range for cybersecurity testing and training, addresses the security of interconnected 5G and IoT networks, including the identification and mitigation of cascading attacks and adversarial AI threats. This makes it crucial for the project to implement a comprehensive risk management strategy, integrating security into the design and development phases and establishing robust incident response protocols. Regular security testing and ensuring the security of third-party components are also essential.

5 The NITRO FAIR Dataset Template Questionnaire

This section presents the NITRO FAIR Questionnaire. The NITRO FAIR Dataset Template Questionnaire (Table 6) includes a set of questions that all Data Controllers are required to fill in for each Dataset [GFDM], [GGD], [ERC].

As mentioned, the DMP along with the information on the following form will be made available gradually through successive updates and newer versions, as the project's implementation progresses. The Data Controllers will be responsible for updating their respective tables every time significant changes occur.

Table 5: NITRO Project's detailed information.

Project Acronym	Project Number
NITRO	101145872
Description	
Title	Name of the Dataset <i>Please provide a meaningful name so we can refer to it unambiguously in the future.</i>
Task	NITRO task/subtask where Dataset was generated <i>Describe the overall setting of the use case in a scenario style, clarify how things will really happen during pilots, who will be involved, who will benefit, etc.</i>
Data owner/controller	Names and addresses of the organisations or people who own/control the Data
Time period covered by the Dataset	Start and end date of the period covered by the Dataset.
Subject	Keywords or phrases describing the subjects or content of the Data
Language	All languages used in the Dataset
Variable list and codebook	All variables in the Data files, with a description of the variable name, length, type, values

Data quality	Description of Data quality standards and procedures to ensure Data quality
File inventory	All files associated with the project, including extensions
File formats	Format of the file
File structure	Organisation of the Data file(s) and layout of the variables, where applicable
Necessary software	Names of any special-purpose software packages required to create, view, analyse, or otherwise use the Data
Details on the procedures for obtaining informed consent	Please provide details on the procedures for obtaining informed consent from the data subjects (e.g., providing an information sheet together with the consent form). In case of children/minors and/or adults unable to give informed consent, indicate the tailored methods used to obtain permission. According to the H2020 Guidelines, if the Data subjects are unable to give consent in writing, for example, because of illiteracy, the non-written consent must be formally documented and independently witnessed. Please explain how you intend to document oral consent. In the very exceptional case that it can't be recorded please give reasons. If you will use deception for another type of Data subject, you must obtain retrospective informed and free consent as well as debrief the participants. Deception requires strong justification and appropriate assessment of the impact and the risk incurred by both researchers and participants.
Measures taken to prevent the risk of enhancing vulnerability/stigmatisation of individuals/groups	<i>Please indicate any such protective measures (e.g. use of anonymisation techniques, use of pseudonyms, non-disclosure of audio-visual materials, voice records, etc.)</i>
Description of the processing operations (i.e. what you do with Personal Data and how)	Processing of ‘Personal Data’ means any operation or set of operations which is performed upon Personal Data, whether or not by automatic means, such as: •Collection (digital audio recording, digital video caption, etc.) •Recording •Organisation and storage (cloud, LAN or WAN servers) •Adaptation or alteration (merging sets, amplification, etc.) •Retrieval and consultation •Use •Disclosure, transmission, dissemination, or otherwise making available (share, exchange, transfer, access to the Data by a third party) •Alignment or combination •Blocking, deleting or destruction, etc. <i>Please describe in detail the processing operations that you will perform to conduct your research and give detailed feedback on participants. Indicate also if a copy of notification/authorisation for tracking or observation is required.</i> <i>Any research activity may involve the processing of Personal Data (ICT research, genetic sample collection, research activities involving personal records (financial, criminal, education, etc.), lifestyle and health</i>

	<i>information, family histories, physical characteristics, gender and ethnic background, location tracking and domicile information, etc.)) any method used for tracking or observing.</i>
--	---

Table 6: NITRO’s FAIR Dataset Template Questionnaire.

1. Data Summary
1.1 Purpose
1.2 Types and Formats of Data
1.3 Re-use of Existing Data
1.4 Origin
1.5 Expected size
1.6 Data utility
2 FAIR Data
2.1 Making Data findable (Dataset description: metadata, persistent and unique identifiers, e.g.)
2.1.1 Are the Data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism?
2.1.2 What naming conventions do you follow?

2.1.3 Will search keywords be provided that optimise possibilities for re-use?
2.1.4 Do you provide clear version numbers?
2.1.5 What metadata will be created?
2.2 Making Data openly Accessible
<i>which Data will be made openly available, and if some Datasets remain closed, the reasons for not giving access; where the Data and associated metadata, documentation and code are deposited (repository?); how the Data can be accessed (are relevant software tools/methods provided)?</i>
2.2.1 Which Data produced and/or used in the project will be made openly available as the default?
2.2.2 How will the Data be made accessible?
2.2.3 What methods or software tools are needed to access the Data?
2.2.4 Is documentation about the software needed to access the Data included?
2.2.5 Is it possible to include the relevant software?

2.2.6 Where will the Data and associated metadata, documentation and code be deposited?
2.2.7 Have you explored appropriate arrangements with the identified repository?
2.2.8 If there are restrictions on use, how will access be provided?
2.2.9 Is there a need for a Data access committee?
2.2.10 Are there well-described conditions for access?
2.2.11 How will the identity of the person accessing the Data be ascertained?
2.3 Making Data Interoperable (<i>which standard or field-specific Data and metadata vocabularies and methods will be used</i>)
2.3.1 Are the Data produced in the project interoperable?
2.3.2 What Data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

2.3.3 Will you be using standard vocabularies for all Data types present in your Data set to allow inter-disciplinary interoperability?
2.3.4 In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
2.4 Increase Data re-use <i>(which Data will remain re-usable and for how long, is embargo foreseen; how the Data is licensed; Data quality assurance procedures)</i>
2.4.1 How will the Data be licensed to permit the widest re-use possible?
2.4.2 When will the Data be made available for re-use?
2.4.3 Are the Data produced and/or used in the project useable by third parties, in particular after the end of the project?
2.4.4 How long is it intended for the data to remain re-usable?
2.4.5 Are Data quality assurance processes described?
3 Allocation of resources
3.1 What are the costs for making Data FAIR in your project?

3.2 How will these be covered?
3.3 Who will be responsible for data management for your project?
3.4 Are the resources for long-term preservation discussed?
4 Data security
4.1 What provisions are in place for Data security?
Please indicate any methods considered for secure Data storage and transfer of sensitive Data.
4.2 Is the Data safely stored in certified repositories for long-term preservation and curation?
5 Ethical aspects / Protection of Personal Data notification of processing operations
5.1 Are there any ethical or legal issues that can have an impact on Data sharing?
5.2 Is informed consent for Data sharing and long-term preservation included in questionnaires dealing with Personal Data?

5.3 Name of the Processor(s) Please indicate the names of any other natural or legal person that may process the Data. If processors can be categorised into groups, please refer to them by groups and not necessarily by name; otherwise, indicate their names.
5.4 Lawfulness of Processing Data Controllers must process only those Personal Data that are necessary during the project and for a specific purpose. Processing Personal Data that are not essential to the research may even expose Data Controllers to allegations of 'hidden objectives', i.e. processing information with the Data subjects' permission for one purpose and then using that information for another purpose without specific permission.
5.5 Categories of Data Subjects Please indicate the categories of Data subjects involved in the processing operations of the project.
5.6 Categories of Personal Data Please list concretely the categories of Personal Data that you will process: • Normal Personal Data: name, home address, e-mail address, location Data, etc. • Sensitive Data: religious beliefs, political opinions, medical Data, sexual identity, etc.

5.7 Rights of Data Subjects

Regarding Article 16 of the EUI's Data Protection Policy, Data subjects enjoy the following rights concerning their Personal Data:

- To be informed whether, how, by whom and for which purpose they are processed
- To ask for their rectification in case they are inaccurate or incomplete
- To demand their erasure in case the processing is unlawful or no longer lawful ('right to be forgotten')
- To block their further processing whilst the conditions under letters b) and c) of this Article are verified.

Note: Please indicate how you will ensure the Data subjects' rights. E.g. participants will be free to withdraw at any time without justification. The Data collected prior to the withdrawal will be deleted. In such a case, you may need to ensure the erasure of the collected Data while maintaining anonymity. In order to do so, you may use a pseudonym for each participant, ensuring that the key to the pseudonyms is password-protected and available only to the Data Controller.

5.8 Safeguards taken to protect the Data subjects' identity.

Regarding Article 2 of the EUI's DP Policy, Identifiable persons can be identified directly or indirectly, in particular by reference to an identification number or to one or more factors specific to their physical, physiological, genetic, mental, economic, cultural or social identity.

Please provide details on the measures taken to avoid direct or indirect identification of the Data subjects, e.g. by using anonymisation techniques or pseudonyms. E.g. names of the Data subjects will not be disclosed, at any time, in audio recording and published material. Pseudonyms (a reversible system of coding in order to be able to re-contact participants if needed) will be used in all documentation, and any additional information that may reveal the identity of participants will be concealed when publishing.

Destroy any residual information that could lead to the identification of participants at the end of the project. You must explain this procedure clearly to participants during the 'recruitment' process.

6 Other issues

6.1 Do you make use of other national/funder/sectorial/departmental procedures for Data management? If yes, which ones?

References

- [ORD] "Open research data and data management plans information for erc grantees," February 2018. [Online]. Available: https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf.
- [GFDM] "Guidelines on FAIR Data Management in Horizon 2020," European Commission/Directorate-General for Research & Innovation, 2016.
- [GDPR] "GDPR EU," [Online]. Available: <https://www.gdpreu.org/the-regulation/key-concepts/data-controllers-and-processors/>.
- [DMEC] "Data management, European Commission, Research and Innovation, Participant Portal H2020 Online Manual," [Online]. Available: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm.
- [ZND] "Zenodo," an interdisciplinary open data repository service maintained by CERN, Geneva. Datasets can be located via the Zenodo search engine., [Online]. Available: <https://www.zenodo.org/>.
- [DSG] "Data sharing guidelines | Cancer Research UK," [Online]. Available: <https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines>.
- [GGD] "Guide on Good Data protection practice, This Guide should be consulted in parallel with the EUI's Data Protection Policy as well as with the EUI Code of Ethics in Academic Research.," [Online]. Available: <https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>.
- [EDP] European Commission(EC), "Ethics and data protection by," 2018. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf.
- [ERC] "ERC Open Research Data Management Plan Template," April 2017. [Online]. Available: <https://erc.europa.eu/content/erc-data-management-plan-template>.
- [DHR] <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- [COE] <https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108>
- [LUS] <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML>
- [EPD] Directive 2002/58/EC of the European Parliament
- [REP] "REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," European Commission, 2018.
- [NGT] NITRO GitLab, www.nitro-gitlab.ds.unipi.gr