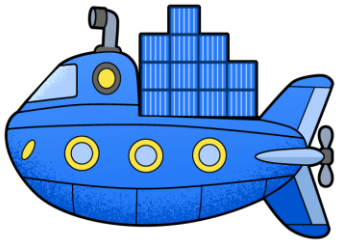

NEMO

Next Generation Meta Operating System

D5.1 Living Labs and data management plan (DMP). Initial version

Document Identification			
Status	Final version	Due Date	28/02/23
Version	1.1	Submission Date	28/02/23

Related WP	WP5	Document Reference	D5.1
Related Deliverable(s)	D1.1	Dissemination Level (*)	PU
Lead Participant	ENG	Lead Author	Antonello Corsi
Contributors	ASM, EMOT, ESOFT, OTE, MAG, NOVO, SU, UPM, SPACE, SIM, INTRA, FHW	Reviewers	Antonis Gonos ESOFT Skias Dimitrios INTRA

Keywords:
GDPR, DMP, Data, Trial site, Living Labs, Open source ,

Document Information

List of Contributors	
Name	Partner
Antonello Corsi	ENG
Francesco Bellesini	EMOT
Maria Belesιoti	OTE
Prashanth PKP Kumar Pedholla	ASM
Antonis Gonos	ESOFT

Document History			
Version	Date	Change editors	Changes
0.1	01/01/2023	ENG	ToC
0.2	10/01/2023	ENG, EMOT, ESOFT, OTE, FHW	Add dataset for trials
0.3	10/02/2023	ENG	First iteration before the drat preparation, added different contributions from different partners
0.4	23/02/2023	ESOFT,ASM,INTRA	Final version with internal reviewers
0.5	23/02/2023	ENG	Ready for quality check
1.0	28/03/2023	ATOS	Quality check and submission
1.1	16/05/2023	ATOS	Re-submission

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Antonello Corsi (ENG)	28/03/2022
Quality manager	Rosana Valle (ATOS)	28/03/2022
Project Coordinator	Enric Pages (ATOS)	28/03/2022
Technical Manager	Harry Skianis (SYN)	28/03/2022

Table of Contents

List of Figures

List of Acronyms

Abbreviation / acronym	Description
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
ISO	International Organization for Standardization
KPI	Key Performance Indicators
LL	Living Lab
MQTT	Message Queuing Telemetry Transport
PMU	Phasor Measurement Unit
PQA	Power Quality Analyzer
SSL	Secure Sockets Layer
WP	Work Package

Executive Summary

The purpose of this document is to outline a data management plan for the NEMO project. Effective data management is critical for the success of any research project, as it helps to ensure that data is well-organized, properly documented, and easily accessible to all members of the research team. This plan outlines the processes and protocols that will be followed in order to manage the data generated by the NEMO project in a responsible and efficient manner. The plan includes details on how data will be collected, organized, stored, and shared with relevant stakeholders. By following this plan, we aim to maximize the impact and value of the data generated by the project.

The main goals of this report are the following:

- To detail the overall methodology for handling the outcomes of the project, in accordance with the H2020 guidelines regarding Open Research Data.
- To list results, information and data that can be published.
- To describe the open repositories for data management and dissemination.

NEMO project partners will provide, through open access, various types of information, such as, scientific publications relevant to the project, white papers published, Open-Source code generated, open datasets, anonymous interview results, etc. It should be stressed that the consortium will balance between open publishing project's related data, collected or generated, and protecting private or sensitive information (according to GDPR provisions) that may have legal implications in case of inappropriate treatment.

The overall analysis towards the definition of NEMO Data Management Plan is provided in this document.

1 Introduction

1.1 Purpose of the document

Based on the guidelines of the Open Research Data Pilot in Horizon Europe[1]the Data Management Strategy for NEMO will be based on the establishment of how data will be handled during and after the project. This process involves:

- A methodology that can make research data generated in the context of the NEMO project: findable, accessible, interoperable and reusable (FAIR principles).
- The identification, classification (i.e., open or confidential), organization, curation, preservation, storing and sharing of the data to be collected, processed and/or generated.
- Requirements related to ethics and legal compliance (i.e., to ensure that the work will be conducted in an ethically sound way) as described in the Grant Agreement and in EU and national legislation.

1.2 Relation to other project work

In today's digital age, data has become the lifeblood of any organization, and managing it effectively can make a significant difference to the success of a project. In general, for each EU funded project where data management is crucial for achieving the desired outcomes a robust DMP is a mandatory prerequisite. It outlines how data will be collected, processed, stored, shared, and preserved throughout the project's lifecycle. Moreover, this plan should be designed in a way that suits the project's specific requirements, and it must be tailored to the specific data generated by the project. In the case of NEMO project, the DMP is originated in T5.1 but covers evenly all the WPs given that this task will initially generate and update an IoT Living Labs Data Management Planning (DMP), collecting data related to IoT devices, CPU and network characteristics and IT equipment (e.g., gateways and firewalls) and their operation under NEMO validation.

1.3 Structure of the document

The deliverable is structured as follows:

- Chapter 1 provides the introduction of the deliverable.
- Chapter 2 presents the NEMO data management strategy, thereby exposing classification, archiving, performance, safety and security, FAIR and ethics requirements and procedures for the data.
- Chapter 3 lists datasets identified at the early phases of the NEMO project.
- Chapter 4 presents recommendations and further steps

[1] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en

2 Data Management Strategy

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon Europe project. As part of making research data findable, accessible, interoperable and re-usable, a DMP should include information about the handling of research data during and after the end of the project [1] :

- What kind of data will be collected, processed and/or generated and to whom might they be useful later on?
- Which methodology and standards will be applied?
- What metadata required to enable data to be found and understood, ideally according to the particular standards of a scientific discipline?
- Whether data will be shared/made open access?
- How data will be preserved (including after the end of the project)?
- How to archive and preserve the open datasets of the project?

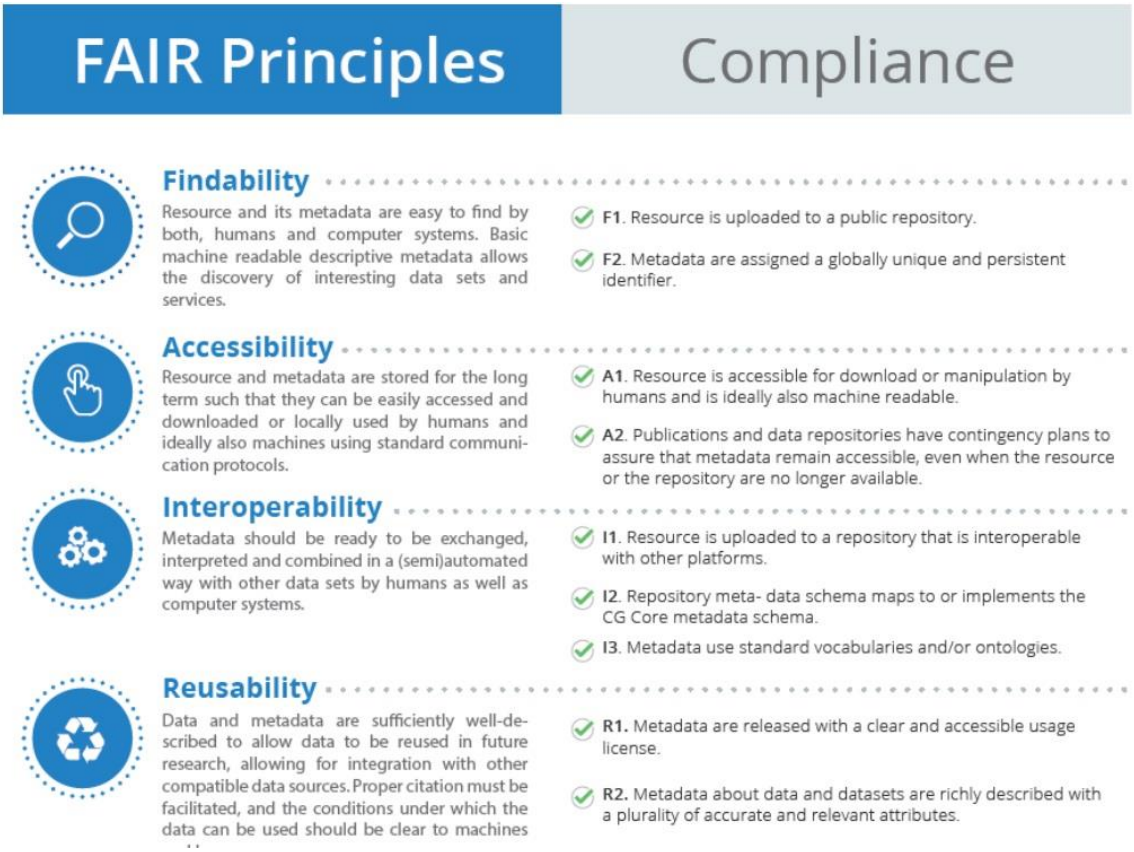

Figure 1: Research Data Management according to FAIR principles[2]

More specifically, for Horizon Europe projects a FAIR DMP template has been designed to be applicable to any project that produces, collect and processes research data. Figure 1 and Figure 2 present the FAIR data principles towards promptly disseminating the data outcomes of a research project.

[2] Image source: CGIAR. Creative commons attribution-non-commercial 4.0 international license.

Data should be **Findable**	F1. (meta)data are assigned a globally unique and persistent identifier (DOI) F2. data are described with rich metadata F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource
Data should be **Accessible**	A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, where necessary A2. metadata are accessible, even when the data are no longer available
Data should be **Interoperable**	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data
Data should be **Reusable**	R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

Figure 2 : Research Data Management according to the FAIR principles data source acquisition[3]

2.1 Data Sources and acquisition

Data collected in NEMO are both public/open data available on the internet and internal data collected or generated from/by partners, mainly LL and research organisations. The data collected in NEMO involves the following data sources:

- Interviews and surveys with stakeholders participating in the pilots during requirements elicitation.
- Public/Open data, such as energy consumption, weather, etc.
- Internal local devices-related data from pilots, such as energy monitoring data, smart meters data, PV plant production, crop images, farm monitoring data, sentiment data, etc.
- Data collected during the project's Open Calls

2.2 Type of data

In addition, a main point of the DMP is the definition of the open access type over the data. Open Access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean:

- peer-reviewed scientific research articles (published in scholarly journals) or
- research data (data underlying publications, curated data and/or raw data).

Open Access to scientific publications means free online access for any user. The two main routes to Open Access are:

[3] Source: WWW.researchgate.net

- Self-archiving / 'green' Open Access – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.
- Open Access publishing / 'gold' Open Access - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors.

Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge. Figure 3 presents the process flow towards defining the open access type in scientific publications and research data.

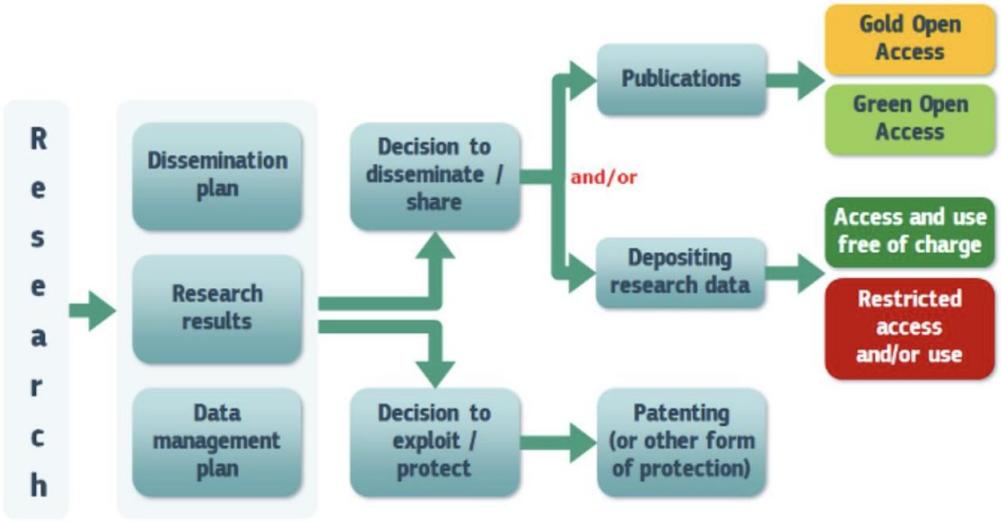

Figure 3: Open Access strategy for publications and research data

The open access mandate comprises two steps:

1) Depositing publications in repositories
2) Providing open access to them

2.3 Data Management Requirements

The NEMO Data Management Process is defined as a management approach for each result generated or collected during the project runtime. Such an approach outlines requirements for several data management aspects, including data classification, data archiving, data performance, data safety and security, FAIR data management and data ethics.

These aspects refer to NEMO platform as well as to other assets of NEMO project, namely NEMO Portal, the project's shared repository (OwnCloud), the research data online repository (Zenodo), NEMO source code repository (GitLab).

The NEMO requirements and corresponding processes for the aforementioned data management aspects are presented in the following paragraphs.

2.3.1 Data Classification

For proper data handling, datasets ought to be primarily classified as public and non-public.

It is important to note that the classification may not always be straightforward and may require careful consideration of the specific circumstances surrounding the data and its intended use. In cases where it is unclear whether a dataset should be classified as public or non-public, general question as the following can help.:

1. Does a result provide significant value to others or is it necessary to understand a scientific conclusion?
If this question is answered with yes, then the result is classified as public (granted for open access). If this question is answered with no, the result is classified as non-public. For example, code that is very specific to the NEMO platform (e.g., a database initialization) is usually of no scientific interest to anyone, nor does it add any significant contribution.
2. Does a result include personal information that is not the author's name?
If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published according to the ethical principles of the project.
3. Does a result allow the identification of individuals even without the name?
This is also a step managed by the ethics management plan of the project as we have committed in NEMO project to establish anonymization techniques to conceal a single user's identity, e.g.: abstraction, dummy users, or non-intersecting features. If this question is answered with yes, the result is classified as non-public.
4. Can a result be abused for a purpose that is undesired by society in general or contradicts with societal norms and the project's ethics?
If this question is answered with yes, the result is classified as non-public.
5. Does a result include business or trade secrets of one or more partners of the project?
If this question is answered with yes, the result is classified as non-public. Business or trade secrets need to be removed in accordance with all partners' requirements before it can be published.
6. Does a result name, technologies that are part of an ongoing, project-related patent application?
If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after the patent has been filed.
7. Does a result break security interests for any project partner?
If this question is answered with yes, the result is classified as non-public.

This is a simple structural approach to determine the different data types defined as part of the DMP. The responsibilities of the NEMO consortium partners towards disseminating the project outcomes are defined in the following section.

2.3.2 Data Archiving

As NEMO technical solution gets more mature, more and more data will be ingested to the platform. Such data can be public-interest data (e.g., weather data), operational data from pilot sites and LL, or personal data from new users. For the first two categories, data will be ingested to the systems at very high resolutions (e.g., hourly). As a result, large volumes of data will be ingested to the system on a daily basis. On the other hand, such data tend to be very useful when they are fresh for the development of real time services, while after a short time period from their ingestion they are mostly used for static analysis and for batch processing analytics. Moreover, the probability to update such data is very low, especially after some days.

Also, it is very expensive to keep such old data on a project's hot storage (Tier-1). At this stage of the project only one data base is used (cold storage) to store project data. However, in this context, NEMO

is evaluating the possibility to create mechanisms to archive old operational and public interest data and retain data to the hot data storage of NEMO only for a specific period of time. The retention period has not been decided yet, but 3 or 6 months are the most popular options. After this period data will be archived to NEMO cold data storage. Data from cold data storage will be available (only read permissions) in batches to users only if they are authorized to have access to it. Of course, access policies for each dataset will be decided by the data owners.

Regarding personal data, only their owners will have access to it (read and write) at any given time. Thus, a user can change his/her personal data at any time. As a result, such data are not suitable for archiving. Only data from deleted users will be examined for archiving, and only for the duration of NEMO project, in order to allow its seamless operation and the correct measurement of project's KPIs. Of course, for users that have deleted their profile, no personal data will be kept. Instead, only aggregate anonymised data will be retained.

2.3.3 Data Performance

As already mentioned in 2.3.2, NEMO will make use of large amounts of different types of data coming from heterogeneous sources and providers. Amongst others, the continuous availability of such data is an indisputable attribute that entails high computational and performance requirements.

NEMO platform envisages a microservice-based architecture that separately addresses each stage of the data flow within the platform i.e., data interoperability and homogenization, data streaming, and data storage. Communication protocols can vary between request-response and publish-subscribe or message queue. Nevertheless, in case of absence of such capabilities from the data serving side, bulk ingestion procedures shall be examined as well (e.g., sftp).

Regarding data interoperability and homogenization, several data connectors are required to automatically ingest the data coming from the different data sources (e.g., high resolution real-time and historical data from sensors and smart meters). The above-mentioned connection protocols and interfaces will be used depending on the nature of each dataset. Data shall be homogenized according to a common data model. To sustain the workloads of such communication for data ingestion and homogenization, scalable cluster architectures shall be examined allowing for distributed and efficient data processing.

Storing ingested and homogenized data within NEMO shall require a high amount of data storage. Distributed file and database systems will be considered, focusing on cloud infrastructure, aiming to achieve high data availability in terms of (a) unstructured ingested data and (b) homogenized data forming the NEMO data lake.

With regards to data streaming, this stage will enable the information exchange amongst NEMO components of the meta-architecture allowing the transition of data to the analytics and modeling phases. All components of the microservice based architecture for data operation will be initially deployed locally.

2.3.4 Data Safety and Security

Most of the datasets that are either collected or generated by the NEMO platform, including any personal data, will be stored in the platform's database repositories. In this way, the project can ensure compliance with the (GDPR-based) "right-to-be-forgotten" requests by the research participants. Furthermore, NEMO will take all necessary measures to assure data security and safety and mitigate any risks associated with the storage and transmission of special categories of personal data. A list of measures is provided that the consortium can adopt according to the advancements and the detailed definition of the technological aspects and pilot scenarios and requirements.

To ensure the data is safe and secure, the following measures can be taken in these regards during the execution of NEMO:

- Assign codes to the organisations in which the research participants whose data is collected during the pilots are participating. In this way it will be impossible to identify the research participant, apart from the partner data controller.
- Adopt decentralised data processes. An active component of the project's research will involve the use of decentralised data systems, in which research participants retain possession of their own personal data, with technological access for project partners solely granted for the purposes of the project.
- Ensure redundancy. To ensure stored data is not lost, fault tolerance principles based on redundancy should be employed. With respect to storage solutions in databases or storage as a service, these translate to performing periodical backs-ups as required by the type of data stored as discussed also in Sections 2.3.2 and 2.4.
- Pseudonymise personal data of research participants. This technique involves changing the data in such a manner that the personal data can no longer be attributed to a specific person without the use of additional information.
- Encrypt personal data of research participants. All personal data should be securely stored in an encrypted form. A key is needed for the decryption of encrypted data. This key should be securely stored in a way that it will be impossible for someone with access to the encrypted data to find.
- Encrypted data sharing connections shall be established allowing safe data exchange through secure channels (e.g. SSL) amongst NEMO microservices and between the platform and external data sources.
- Only authorised and hence authenticated researchers from partner organisations of the NEMO project consortium shall have access to the personal data of the research participants.
- Passwords used should be strong enough. Strong passwords will be enforced by the configurations in all NEMO systems. All relevant technological components such as Istio, Apache NIFI, MongoDB, etc, will enforce strong passwords.
- Activities shall be monitored and logged allowing for traceability and latter data regulation compliance assessment. Logging will be enabled in the specific tools and components including MongoDB, InfluxDB, MinIo, Istio, etc.
- System and software vulnerabilities shall be detected and mitigated enabling the protection from unauthorised access and data breaches.
- Microservice configurations and data shall be regularly backed up externally allowing for data persistence and harmonised functionality in case of failures or crashes. Configuration and data back-up and archiving are realised.

The project will preserve the data collected and generated during the lifetime of the project in order to allow the achievement of the project´s vision and goals. All personal data collected will be deleted at the end of the project. The rest of the research datasets will be shared as open data using the Linked Open Data standards and open licenses according to the Open Research Data Pilot in Horizon Europe.

2.3.5 FAIR Data

NEMO project supports the reuse of research data and follows FAIR principles [2], [3]. FAIR represents a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable.

- **Findable:** data has a unique, persistent ID, located in a searchable resource, and documented with meaningful metadata.
- **Accessible:** data is readily and freely retrievable using common methods and protocols, metadata is accessible even if the data is not.

- **Interoperable:** data is presented in broadly recognised standard formats, vocabularies, and languages.
- **Re-useable:** data has clear licences, and accurate meaningful metadata conformity to relevant community standards and identifying its content and provenance.

2.3.5.1 Findable DATA

The data that will be generated in the framework of NEMO shall be easily discovered by research communities. For this reason, these data will have a unique identifier. The suggested identifier is:

NEMO_[Name]_[Type]_[Date]_[Owner]

Where

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

For example, the following identifier

NEMO_PlatformConfiguration_code_30012023_OTE

specifies that the generated dataset is a code which is used to configure the platform, and that the configuration is completed on 30-01-2023 by OTE, which is the owner of the configuration. Apart from the unique name identifier, the data set description is also important. It is essentially an expanded description of the identifier with more details. Additionally, it is important to clarify which data are generated by the project and which are collected and used.

The data set description is organized as the metadata. Metadata is used to give information about other data. The metadata identifier is formed in the similar way as the data identifier but with more details and, depending on the file format, will be either incorporated as a part of the file or as a separate file (in its simplest form) in the text format. The suggested identifier is:

NEMO_[Name]_[Type]_[Date]_[Owner]_METADATA

Where

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

Additionally, all assets will be tagged with a Digital Object Identifier (DOI), a persistent identifier or handle used to identify assets uniquely, standardized by the International Organization for Standardization (ISO). DOIs are pragmatically useful for sharing, accessing, searching, citing, and relating assets.

2.3.5.2 Data Privacy and Ethics

The NEMO project consortium conforms to the regulations of the user informed consent and privacy policies and regulations of the EU. Consortium will act according the:

- General Data Protection Regulation 2016/679 (Protection of personal data).

- Opinion 23/05/2000 of the European Group on Ethics in Science and New Technologies concerning

'Citizens' Rights and New Technologies: A European Challenge' and specifically those relating to:

- ICT (Protection of privacy and protection against personal intrusion).
- Ethics of responsibility (Right to information security).
- Article 15 (Freedom of expression and research and data protection).

The project will ensure that the Consortium Agreement (or addendums thereof) is constructed to enable such assurances to be formally made and adhered to by consortium partners. In addition, with respect to General Data Protection Regulation 2016/679 (Protection of personal of data), individual work packages will be specifically requested to ensure that any models, specifications, procedures or products also enable the project end users to be compliant with this regulation.

The NEMO partners also will abide by professional ethical practices and comply with the Charter of Fundamental Rights of the European Union[4]

The personal datasets identified so far are the following:

- Personal data for coordination and administration purposes: This includes data of the project members related to administrative and general management issues in the framework of the project.
- Personal data for dissemination purposes coming from the web site: This includes webpage cookies, name and e-mail of subscribers to newsletter and lists, contact of people involved in events, etc.

2.3.5.3 IPR and knowledge management plan

The Consortium Agreement includes all the information and rules that project partners take under consideration in order to define the important points necessary to obtain the best possible management (financial conditions, Intellectual Property Rights (IPR), planning) of intellectual property. IPR will be managed in line with a principle of equality of all the partners towards the foreground knowledge and in full compliance with the general Commission policies regarding:

- Ownership.
- Exploitation rights.
- Confidentiality.

In general, outcomes, innovative ideas, concepts and solutions that are not going to be protected by patent applications by the partners will be made public after agreement between the partners, to allow others to benefit from these results and exploit them. However, where results require patents to show the impact of NEMO, there will be thorough scrutiny to determine that this does not infringe on patents belonging to others. Additionally, the intellectual property rights belonging to third parties will be considered and consortium members will ensure no infringement on intellectual property rights. The unified Consortium Agreement will be used as a reference for all IPR cases. The Consortium Agreement identifies the background intellectual property of each of the partners that may be used to achieve the project objectives.

The Consortium Agreement will provide rules for handling confidentiality and IPR to the benefit for the Consortium and its partners. All the project documentation will be stored electronically in the project's

[4] https://www.europarl.europa.eu/charter/pdf/text_en.pdf

website and a repository. Classified documents will be handled according to proper rules with regard to classification, numbering and locked storing and distribution limitations.

In all cases, any action or decision made by the consortium will be subject to the provisions in the CA. The policy, that will govern the IPR management in the scope of NEMO, is driven by a set of principles described in the following sections.

a) Ownership of outcomes

Each partner who produces an outcome under NEMO project will also be the owner of this outcome. If more than one partners take part in the outcome production in a way that could not be possible to separate the contribution of each partner, then all the involved partners will have joint ownership of this outcome.

b) Protection of outcomes

Outcomes protection is not mandatory in all cases. However, the valuable outcomes should be protected. If more than one partners own an outcome, then all these partners should agree if the outcome will not be protected. Where an outcome is capable of industrial or commercial application, it should be protected in an adequate and effective manner in conformity with the relevant legal provisions. The owner or the owners should provide the best strategy to protect the outcome in order to become a commercial product. Although a partner that owns an outcome is not required to consult the rest of the partners in order to decide to protect or not the outcome, they should preferably be informed, because they could express possible legitimate interests. If an outcome owner is not willing to protect it or extend its protection, then EC should be informed at least 30 days prior to the corresponding deadline. EC could protect this outcome if it is considered to be necessary, with the agreement of the contractor concerned.

c) Use and dissemination

The provisions for use and dissemination are the following:

- The partners shall use or cause to be used the outcomes arising from the project, which they own, in accordance with their interests. The contractors shall set out the terms of use in a detailed and verifiable manner, notably in the plan for using and disseminating the outcomes.
- If dissemination of outcomes would not adversely affect its protection or its use, the contractors shall ensure that it is disseminated within a period of two years after the end of the project.

d) Access rights

The access rights for implementation are the following:

- Project partners shall enjoy access rights to the outcome and to the pre-existing know-how, if that outcome or pre-existing know-how is needed to carry out their own work under that project. Access rights to outcome shall be granted on a royalty-free basis. Access rights to pre-existing know-how shall be granted on a royalty-free basis, unless otherwise agreed before signature of the contract.
- Subject to its legitimate interests, the termination of the participation of a project partner shall in no way affect its obligation to grant access rights pursuant to the previous paragraph to the other contractors until the end of the project.

The access rights for outcome use are the following:

Partners shall enjoy access rights to outcome and to the pre-existing know-how, if that outcome or pre-existing know-how is needed to use their own outcome. Access rights to outcome shall be granted on a

royalty-free basis, unless otherwise agreed before signature of the contract. Access rights to pre-existing know-how shall be granted under fair and non-discriminatory conditions to be agreed.

Subject to the partner's legitimate interests, access rights may be requested under the conditions laid down in the previous paragraph until two years after the end of the project or after the termination of the participation of a partner, whichever falls earlier, unless the partners concerned agree on a longer period. The Consortium Agreement ensures the following aspects:

- Confidentiality.
- Outcomes ownership / joint ownership of outcomes.
- Legal protection of results (patent rights).
- Commercial exploitation of results and any necessary access right.
- Commercial obligation.
- Relevant Patents, know-how, and information Sublicense.
- Pre-existing know-how excluded from contract.

Nevertheless, many specific IPR cases, that will need a concrete solution from the bases previously fixed, may exist. In these conflict situations, the Project Management Board will be the responsible to arbitrate a solution. In case of any members of this Board is directly affected by the conflict, it will not participate in the arbitration process.

The NEMO project partners use this plan as a reference for data management (naming, providing metadata, storing and archiving) within the project each time new project data is produced. Relevant questions from partners will also be addressed within WP1. The DMP will not be used as a fixed document. It will be updated from its creation to the end of NEMO project.

2.3.5.4 Making data interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

The expected type of data to be handled within the NEMO project is research data as reports or spreadsheets, test data (simulated or historical anonymized data) as spreadsheets and source code that might be subject to license and/or copyright.

- Reports. Research data will be consolidated in specific reports, i.e. deliverables and will be shared with the research community via open repositories and project website. They will be indexed with metadata as described in previous sections.
- Test data. Generated and/or measured data together with the resulting data will follow standards or methodologies and will be stored and if appropriate shared in interoperable formats like: JSON, XML, REST, CSV.
- Source code. Whenever possible, the NEMO project will adhere to standards for formats, as much as possible compliant with available (open) software applications.

2.3.5.5 Increase data re-use (through clarifying licenses)

Relevant research output will be timely published in scientific journals and/or in relevant conference proceedings and will be made available in open repositories. NEMO project will take measures to enable third party access to this research data. Creative Commons license is considered for this. Data generated in the NEMO project under the public reports is intended to be reusable not only for the duration of the project lifecycle but for as long as they are relevant.

2.3.5.6 Anonymization and pseudonymization

Processing of personal data for NEMO scientific and research purposes will take place in accordance with the GDPR, hence provided that appropriate technical and organizational measures are implemented (GDPR Article 5(e)). Such measures may include anonymization/pseudonymization techniques (GDPR Article 89).

Data anonymization refers to its processing with the aim of irreversibly preventing the identification of the individual to whom it relates. If anonymization is not feasible, data pseudonymization should be examined. Data pseudonymization corresponds to replacing any identifying characteristics of data with a pseudonym, i.e., a value that does not allow the data subject to be directly identified. Pseudonymization should be distinguished from anonymization, as it only provides a limited protection for the identity of data subjects, as it still allows identification using indirect means.

Regarding the anonymization of tabular datasets, that are the main data product of NEMO LL and applications, there are four main anonymization techniques [4] namely:

- k-anonymity privacy model [5]
- ℓ-diversity privacy model [6]
- t-closeness privacy model [7]
- Differential privacy model [8]

Anonymization techniques aim to modify the original values of a table in such a way that user's privacy can be protected while retaining significant utility in anonymous data for the analysis and modelling.

To enable this transformation, and depending on the selected privacy model, there are various anonymization operations available for privacy preserving data, suppression, generalization, data swapping, data perturbation and synthetic data and that are going to be examined by the project. Data to be utilized within the context of NEMO will be already completely anonymized before publication or sharing to the consortium and will not include any personal or sensitive data. The de-identification of datasets has to occur before the beginning of the ingestion: NEMO datasets have to be stripped of any direct identifiers (suppression / data masking) or use synthetic data in a way that eliminates the risk of re-identifying the sensitive data. At the same time, efforts have to be done to minimize possible link ability and actual linkages in order to reduce the risk of data breach and safeguard the anonymity of datasets.

In order to provide the ability, if desired and applicable, to share data sets (e.g., hackathons) without private information anonymization tools can be used during the data ingestion process to protect such information by complete data removal-suppression, generalization or pseudonymity.

2.4 Data archiving and preserving infrastructure

Along with the definition of the datasets, special focus is delivered at the selection of the platform to archive and preserve the datasets. When we choose a repository, it is important to consider factors such as whether the repository:

- Gives the submitted dataset a persistent and unique identifier. This is essential for sustainable citations – both for data and publications – and to make sure that research outputs in disparate repositories can be linked back to particular researchers and grants.
- Provides a landing page for each dataset, with metadata that helps others find it, tell what it is, relate it to publications, and cite it. This makes research more visible and stimulates the reuse of the data.

- Helps to track how the data has been used by providing access and download statistics.
- Responds to community needs and is preferably certified as a 'trustworthy data repository', with an explicit ambition to keep the data available in the long term.
- Matches particular data needs (e.g.: formats accepted; access, back-up and recovery, and sustainability of the service). Most of this information should be contained within the data repository's policy pages.
- Provides guidance on how to cite the data that has been deposited.

3 Living Lab trials description

In this section, we will provide an introduction to the infrastructure of each trial, along with a list of all the datasets that will be relevant to them.

3.1 Trial #1 NEMO Integration Infrastructure Technology Lab

3.1.1 Trial description

The trial #1 is the infrastructure for the technology lab. The datasets that will be used here are the ones indicated in the other trials.

3.2 Trial #2: Smart Farming Use Case & Living Lab

3.2.1 Trial description

This living lab will combine micro-climate/soil/leaf sensing stations (Synelixis' SynField©[5])oT devices), agri-drones and mobile robots to provide a complete solution towards olive fly precise insecticide bio-spraying. The procedure will be autonomous and highly configurable, allowing the farmer to customize it based on their current needs. The pilot will validate the NEMO framework towards achieving precision aerial and terrestrial spraying through cost-, energy- and computationally efficient processes, which remain transparent to the farmer. The expected outcomes of the pilot validation include the cost and environmental benefits arising from optimized (reduced) use of insecticides, as well as increased quantity and quality of produced olives and olive oil, preserving the organic certification of olive oil.

Datasets	
Dataset #1	**Aerial olive tree image dataset.**
Brief Description	This dataset contains aerial images of olive groves, including images of trees (aerial leaf images) and control images (random images taken over an olive grove, but do not contain olive trees).
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The photos will be taken by the aerial drone using a multi-spectral camera. Regarding ethical considerations, the images should not have persons or signs that indicate the location of the field.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	The dataset will be stored in a folder containing two subfolders which will have olive trees and control images respectively.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the classes of the image dataset and the intended use cases. Interoperability issues will be considered for (meta)data.

[5] https://www.synfield.gr/

Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	The dataset validation, which will ensure, e.g., that olive images correctly show olive trees, will be done manually by our team
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The dataset is planned to be publicly available in popular open repositories, e.g., Zenodo, to encourage researchers to develop cutting edge classification models and increase the impact of the project.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The dataset will be managed, according to ISO/IEC 27001 standard. Long-term access will be available through the public repository, in which data will be uploaded.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The dataset will be managed, according to ISO/IEC 27001 standard. Data will be stored on encrypted disks; computer accounts will be password protected and communication security measures will be applied.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	Data governance procedures will be compliant with ISO/IEC 27001 and GDPR.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The DMP is compliant with national (Greek) and EU laws and the GDPR.
Dataset #2	**Field micro-climate dataset**
Brief Description	This dataset contains tabular data collected from SynField IoT devices and integrated sensors. Micro-climate data (e.g., air temperature, air humidity, wind direction, wind speed, rain volume, rain intensity), soil and olive tree related data (leaf wetness, soil type, soil temperature, soil humidity, soil conductivity). The data will be associated with time information and geospatial/location information provided by GPS.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data will be collected through IoT devices placed in the olive grove. No ethical issues arise from the climate/soil/leaf sensor measurements collected.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	The dataset will be stored in tabular format as json or csv files.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the classes of the field dataset and the intended use cases. Interoperability issues will be considered for (meta)data.

Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	Data validation and cleaning techniques are in place for SynField data.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The dataset is planned to be publicly available in popular open repositories, e.g., Zenodo, to encourage researchers to develop cutting edge classification models and increase the impact of the project.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The dataset will be managed, according to ISO/IEC 27001 standard. Long-term access will be available through the public repository, in which data will be uploaded.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The dataset will be managed, according to ISO/IEC 27001 standard. Data will be stored on encrypted disks; computer accounts will be password protected and communication security measures will be applied.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	Data governance procedures will be compliant with ISO/IEC 27001 and GDPR.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The DMP is compliant with national (Greek) and EU laws and the GDPR.
Dataset #3	**Terrestrial olive tree image dataset.**
Brief Description	This dataset contains images taken by terrestrial robots in olive groves.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The photos will be taken by terrestrial robots using a stereo camera. Regarding ethical considerations, the images should not have persons or signs that indicate the location of the field.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	The dataset will be stored in a folder containing two subfolders which will have olive grove and control (ground truth) images.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the classes of the dataset and the intended use cases. Interoperability issues will be considered for (meta)data.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	The dataset validation will be done manually by our team.

Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The dataset is planned to be publicly available in popular open repositories, e.g., Zenodo, to encourage researchers to develop cutting edge classification models and increase the impact of the project.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The dataset will be managed, according to ISO/IEC 27001 standard. Long-term access will be available through the public repository, in which data will be uploaded.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The dataset will be managed, according to ISO/IEC 27001 standard. Data will be stored on encrypted disks; computer accounts will be password protected and communication security measures will be applied.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	Data governance procedures will be compliant with ISO/IEC 27001 and GDPR.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The DMP is compliant with national (Greek) and EU laws and the GDPR.

3.3 Trial #3 Smart Energy & Smart Mobility/City Use Cases & Living Lab

3.3.1 Trial description

ASM acts as distribution system operator, supports the local municipality for e-mobility plans, and includes Umbria Energy as group supplier/retailer. EMOT acts as charging point operator, electric mobility service provider and electric vehicle aggregator. In NEMO ASM and EMOT supported by ENG and TSG, will focus on improving the ability to monitor the smart electricity grid and stabilize it by improving Renewable Energy Sources load balancing via EV chargers, driver-friendly scenarios for smart city mobility and dispatchable charging of EVs based on RES demand-response along with human-centred smart micro-contracts and micro-payments.

Datasets	
Dataset #1	**Smart Meter eXtension (SMX) data**
Brief Description	Measurements of voltage, currents, and power parameters of ASM energy units
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data will be transmitted through MQTT protocol via public IP and / or LAN/VPN connections.

Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Every Smart Meter has a Unique ID and the historical data will be stored in an SQL database.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the dataset.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Data will be shared with all Project Consortium, any usage other than NEMO Project must be agreed with ASM
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Data are stored in a AVEVA Historian Database for 5 years after the end of the projects
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Data is anonymized, access to the database is allowed only to specific IPs and with a password, access to MQTT Broker is allowed with a password.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM is responsible for the maintenance of the database and the Data protection officer of the organization ensures compliance with the EU data protection law
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data stored and the ASM internal management plan follows all the national and European laws in this topic.
Dataset #2	**Smart Meter eXtension (SMX) data**
Brief Description	Measurement of electric power signals to determine the load's ability to function properly with that electric power.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data will be transmitted through MQTT protocol via public IP and / or LAN/VPN connections.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	PQA's have Unique ID and the historical data will be stored in an SQL database.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the dataset.

Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Data will be shared with all Project Consortium, any usage other than NEMO Project must be agreed with ASM
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Data are stored in a AVEVA Historian Database for 5 years after the end of the projects
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Data is anonymized, access to the database is allowed only to specific IPs and with a password, access to MQTT Broker is allowed with a password.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM is responsible for the maintenance of the database and the Data protection officer of the organization ensures compliance with the EU data protection law.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data stored and the ASM internal management plan follows all the national and European laws in this topic.
Dataset #3	**Phaser Measurement Units (PMU) data**
Brief Description	Measurement of Phasors, to improve operators' visibility into what is happening throughout the vast grid network for identifying and analyzing system conditions.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data will be transmitted through MQTT protocol via public IP and / or LAN/VPN connections.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	PMU has a Unique ID, and the historical data will be stored in an SQL database.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the dataset.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Data will be shared with all Project Consortium, any usage other than NEMO Project must be agreed with ASM
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Data are stored in a AVEVA Historian Database for 5 years after the end of the projects

Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Data is anonymized, access to the database is allowed only to specific IPs and with a password, access to MQTT Broker is allowed with a password.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	ASM is responsible for the maintenance of the database and the Data protection officer of the organization ensures compliance with the EU data protection law.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data stored and the ASM internal management plan follows all the national and European laws in this topic.
Dataset #4	**Charging station data**
Brief Description	Charging station voltage, current and power
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data is collected each 1 second using OCPP protocol, an application protocol for communication between charging stations and EMOT central management system, and websocket, a computer communications protocol, providing full-duplex communication channels over a single TCP connection. Data collected are not personal data.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Every charging station has a Unique ID and the historical data will be stored in an SQL database
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the dataset.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Data is shared with technical partners using MQTT. Data must not be used outside the project, for other purpose and/or by Entities which are not part of the consortium
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Data is stored in EMOT server; data will be kept for another five years after the end of the project.
Security and confidentiality: measures in place to protect the data and maintain confidentiality,	Data is anonymized, access to the database is allowed only to specific IPs and with a

including data encryption, access controls, and data use agreements.	password, access to MQTT Broker is allowed with a password.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	EMOT is the owner of charging station dataset
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data stored and the ASM internal management plan follows all the national and European laws in this topic.
Dataset #5	**Electric Vehicle data**
Brief Description	Electric Vehicle State of Charge (SoC), odometer and speed
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	Data is collected each 5 seconds using TCP/IP protocol and MQTT. Data collected are not personal data.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	Every electric vehicle has a Unique ID and the historical data will be stored in an SQL database
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	There will be documentation regarding the dataset.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	A quality indicator is available for the data.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Data is shared with technical partners using MQTT. Data must not be used outside the project, for other purpose and/or by Entities which are not part of the consortium
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	Data is stored in EMOT server; data will be kept for another five years after the end of the project.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	Data is anonymized, access to the database is allowed only to specific IPs and with a password, access to MQTT Broker is allowed with a password.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	EMOT is the owner of charging station dataset
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The data stored and the ASM internal management plan follows all the national and European laws in this topic.

3.4 Trial #4 Smart Manufacturing & Industry 4.0 Use Cases & Living Lab

3.4.1 Trial description

This use case targets ADAS manufacturing. Today, handling and transport of material (SMD-Components) from Auto Store to production sites is performed manually every 30 minutes. By utilizing a 3D-Vision-Camera for Bin Picking Application, integrated Barcode Scanner and collaboration between different robot systems (one cobot and several types of AGVs), we aim to fully automate controlled material picking from Auto Store and autonomous transfer to the production line.

Datasets	
Dataset #1	
Brief Description	Improve mass production and safety in factories with high levels of automation, enabling Collaborative Robot (Cobots) systems, Automated Guided Vehicles (AGVs) and human's co-work.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	material number and material amount are provided from logistic and production planning
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	the material requirement and the material ID is reported to the warehouse depending on the production quantity and the current consumption
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	– material ID – lot number – material amount
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	data comparison: scanned material code with specification (true or false)
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	– no data sharing – data distribution: SARA
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	SAP-documentation and logistic stock list
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	protected by internal network. Access restricted for individual persons

Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	Access restricted for individual persons
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	no compliance issues

3.5 Trial #5 Smart Media City & XR Use Cases & Living Lab

3.5.1 Trial description

Smart Media/City

The round of Athens case: Test boundaries of live media capture and user involvement

During the race, media content is captured by many spectators along the running circuit using smartphones, a few professional and CCTV cameras, and drones.

Incoming content is automatically processed, annotated and rendered (partially on the device using already trained AI/ML models and partially at the edge), and a selection is directly broadcasted (e.g., via social media) based on location info of the (top) runners and interesting events during the race (e.g., based on contributor annotation).

Main objectives:

Investigate advanced networking and data processing NEMO features, such as:

- At AI IoT level, through AV and image analysis using ML techniques, calculate location of a user and multi-users' relevant movement to personalize the view/sound. Do speech/Gesture recognition.
- At Edge cloud, for automated processing and rendering of virtual and augmented objects.
- At cloud level to enable true shared experience even with remotely located users

Datasets	
Dataset #1	Round of Athens
Brief Description	Media content during the race: GPS Data, Video/Image from image/video devices (cameras, smartphones), official information from the race (demographics runners, numbers, etc.), weather, climate/CO_2 (open data)

Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The AV content will be acquired from different audiovisual devices able to record and transmit audio, video and linked metadata such as professional video cameras, drones and personal devices that provide user-generated content. GPS data embedded with the media coming different devices will be used to geolocate the content and link it to additional augmented content. Other open data from the city of Athens will be gathered and used to enrich the content. Content obtained on public roads and from marathon runners must have the necessary permits. Filming with drones may require special permissions (need to be checked).
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	The material will be described by content, source, origin, format and technical parameters, all these gathered in a single unique id that will identify the content. Additionally, the metadata related to the video will be structured following EBU models. The data coming from open data sources will be kept as they were acquired.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	The material will be described by content, source, origin, format and technical parameters, all these gathered in a single unique id that will identify the content. Additionally, the metadata related to the video will be structured following EBU models. The data coming from open data sources will be kept as they were acquired.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	All AV content sources will be filtered according to a quality threshold (QoE analysis) For user generated content, an automated analysis to eliminate content not suitable for all audiences will be done.
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	Some content pieces will be created from the produced material for the Project dissemination and social media purposes. No open remote Access will be provided to the final content.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	As the pilot will deal with live real-time video transmission, most of the video material will not be recorded and stored. For the final video program and other data sources a repository will be created.

	This will allocate the audiovisual sources, like the tourism information, and other metadata gathered from Athen's open data. The final produced content (the one finally delivered and the content pieces) will be archived in a private repository and a long-term hard copy (LTO).
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	No open remote Access will be provided to the final content.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	The produced content will be delivered using a content delivery network. All the material used to create it will be exclusively for internal use of the Project and will not be available.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	Content obtained on public roads and from marathon runners must have the necessary permits. Filming with drones may require special permissions (need to be checked)

XR Use Cases

VR Experience about ancient Workshop of sculptor Phidias

The visitors of the premises of the FHW participating to an VR application that presents everyday life in the workshop of a famous sculptor, the users are able to live the experience of participating in several actions and activities as they are executed in the workshop. The XR application will be based on heterogeneous IoT devices (i.e., Wearables, AR/VR headsets etc.) and is going to collect and analyze biometric data from the users in order to estimate their emotional and physical status during the VR experience. The specific use case is going to use state-of-the-art machine learning algorithms that are going to be trained and executed in the IoT-to-Edge-to-Cloud continuum.

Enhance AV experience in the Tholos Dome VR Theatre

This scenario is going to further enhance the user's audiovisual (AV) experience in Tholos by providing the appropriate software and tools to support multi-sensorial stimuli via effects such as wind, heat, and vibration, or movement and sound in addition to audiovisual (AV). This scenario will analyze the physical position of the presenter as well as his/her speech and it will perform gesture and voice recognition based on state-of-the-art machine learning algorithms. The system trains and executes ML models in the IoT-to-Edge-to-Cloud continuum and it will trigger events in real time that are going to be consumed by multi-sensorial stimuli systems, AR/VR glasses, wearables, and smart displays outside the Tholos.

This use case aims to combine data from heterogenous devices such as XR headset and smart watches to create an interactive augmented reality environment.

Datasets	
Dataset #1	**Emotion recognition multimodal dataset with voice and smart watch data.**
Brief Description	This dataset will combine data originating from 2 different sources, namely heart rate from smart watches and voice from microphones. These instances will be measured concurrently, and label based on the corresponding emotion.
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The methods used to collect the data will be smart watches and microphones attached to the participants at the same time. The dataset will not contain sensitive information such as name or address.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	The dataset will be used in a folder containing 2 subfolders, each for a different data source. The corresponding names will be smart_watch_dataset and voice_dataset.
Data documentation: how the data will be described and documented, including metadata standards and data dictionaries.	Each of the 2 parts of the dataset will have a data dictionary explaining the datatype of each feature and a short description describing what this feature means.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	To ensure data quality, firstly the tools must be frequently checked to ensure they are working properly. Secondly, a static validation will be performed to ensure no missing vales and certain conditions are met (e.g., length of voice input is more than a specific threshold).
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The dataset will be hosted in a public repository to allow public access by researchers and practitioners. A specific license will be chosen that allows the use and distribution of this dataset freely.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The dataset will also be stored in a backup cloud server inaccessible by the public to ensure preservation. Also, there will be a detailed data migration plan to correctly transfer and setup this database in a different server.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The dataset will be made public so confidentiality and security regarding the stored attributes will not be an issue. However, sensitive features will be encrypted.
Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	The database administrator will be responsible for the availability and the malfunction repair of the server and the database. Moreover, a specific group of developers will have access and will be

	responsible for extending the database with possibly new instances.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The dataset will comply with EU relevant laws and the GDPR.
Dataset #2	**Hand gesture recognition dataset.**
Brief Description	This dataset will contain images of hand gestures and a control group with relevant labels
Data collection: methods used to collect the data, instruments or tools used, and any ethical considerations.	The methods used to collect the data will be a camera setup in the same environment that the inference will take place. The dataset will not contain sensitive information such as faces.
Data organization: how the data will be organized, including file naming conventions, data structure, and data dictionaries.	The dataset will be used in a folder containing N subfolders, each for a different class. Each class will have specific documentation that will explain the specific gesture and thus the choice of label.
Data quality control: methods used to ensure data quality, including data validation procedures and data cleaning techniques.	To ensure data quality, firstly the tools must be frequently checked to ensure they are working properly. Secondly, a static validation will be performed to ensure no missing values and certain conditions are met (e.g., images contain a range of pixel values).
Data sharing and dissemination: how the data will be shared and distributed, including data sharing agreements, access and use restrictions, and data deposition plans.	The dataset will be hosted in a public repository to allow public access by researchers and practitioners. A specific license will be chosen that allows the use and distribution of this dataset freely.
Data preservation and archiving: how the data will be preserved and archived for long-term access, including backup and storage strategies, data migration plans, and preservation metadata.	The dataset will also be stored in a backup cloud server inaccessible by the public to ensure preservation. Also, there will be a detailed data migration plan to correctly transfer and setup this database in a different server.
Security and confidentiality: measures in place to protect the data and maintain confidentiality, including data encryption, access controls, and data use agreements.	The dataset will be made public so confidentiality and security regarding the stored attributes will not be an issue. However, sensitive features will be encrypted.

Data governance: roles and responsibilities of individuals and groups involved in managing the data, including data stewardship, data management plan oversight, and data governance policies.	The database administrator will be responsible for the availability and the malfunction repair of the server and the database. Moreover, a specific group of developers will have access and will be responsible for extending the database with possibly new instances.
Compliance: how the data management plan aligns with relevant laws, regulations, and institutional policies, including data privacy and security regulations.	The dataset will comply with EU relevant laws and the GDPR.

3.6 Trial #6 NEMO multi Living Labs Federation

3.6.1 Trial description

Trial 6 will validate cross-living scenarios and demonstrate the applicability of the NEMO meta-architecture and meta-OS in diverse environments and application domains. At this point of the project, issues related with heterogeneous fog/IoT/5G networks, federated learning training and data sharing across LLs are expected to be validated with no additional datasets compared to the ones that will become available from trials 1 to 5.

4 Recommendations and further steps

This deliverable has been dedicated to the first iteration of the DMP for the NEMO Project. As first version of the data management plan (DMP) this document has outlined the way in which datasets in the different trials will be generated, organized, documented, stored, backed up, preserved, and shared, if possible, with other researchers after the publication of the project's main results. In the next version of the deliverable the DMP should identify, and designate, key responsibilities for data management within the project's consortium. The individual typically designated as the principal investigator is ultimately accountable for managing the data for the project. Following this approach in the next version of this deliverable a Data Manager Officer will be identified for each pilot for overseeing data gathering and quality management and creating/executing the DMP of each pilot site.

The DMP in this sense it is a living document that is continuously shared and updated by the overall consortium project team and reviewed regularly when necessary.

5 References

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

[2] Guidelines on FAIR Data Management in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[3] FAIR data, https://en.wikipedia.org/wiki/FAIR_data.

[4] A. Majeed, S. Lee, "All Authors Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey", IEEE Access, Vol. 9, pp. 8512-8545, Dec. 2020.

[5] L. Sweeney, ''k-anonymity: A model for protecting privacy,'' Int. J. Uncertainty, Fuzziness Knowl.-Based Syst., vol. 10, no. 05, pp. 557–570, Oct. 2002.

[6] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, ''l-diversity: Privacy beyond k-anonymity,'' ACM Trans. Knowl. Discovery From Data, vol. 1, no. 1, p. 3, 2007.

[7] N. Li, T. Li, and S. Venkatasubramanian, ''T-closeness: Privacy beyond k-Anonymity and l-Diversity,'' in Proc. IEEE 23rd Int. Conf. Data Eng., Apr. 2007, pp. 106–115.

[8] C. Dwork, ''Differential privacy: A survey of results,'' in Proc. Int. Conf. Theory Appl. Models Comput. Changsha, China: Springer, 2008, pp. 1–19.