

MANOLO

CLOUD-EDGE EFFICIENT & TRUSTWORTHY AI

D8.3 Data Management Plan v.2

Q-PLAN INTERNATIONAL

30/06/2025

Document Information

Issued by:	Q-PLAN INTERNATIONAL
Issue date:	30/06/2025
Due date:	30/06/2025
Work package leader:	NUIDUCD - CeADAR
Dissemination level:	PU-Public

Document History

Version	Date	Modifications made by
0.1	07/05/2025	Q-PLAN drafted the first version and circulated templates for input from partners
0.2	06/06/2025	Full draft version shared for Quality Review
1.0	30/06/2025	Final version ready for submission

Authors

Name(s)	Beneficiary
Tsolakis A., Christodoulidis P., Spyridopoulos G., Theodoridou T.	Q-PLAN

In case you want any additional information, or you want to consult with the authors of this document, please send your inquiries to: spyridopoulos@qplan-intl.gr

Quality Reviewers

First Name	Beneficiary
Cristian Bosch Serrano, Ricardo Simon Carbajo	NUIDUCD - CeADAR

Disclaimer

Funded by the European Union under GA no. 101135782. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or CNECT. Neither the European Union nor the granting authority can be held responsible for them.

©MANOLO Consortium, 2025

Reproduction is authorised provided the source is acknowledged.

Table of Contents

EXECUTIVE SUMMARY	5
1. INTRODUCTION	6
1.1 OBJECTIVES	6
2. DATA SUMMARY	9
2.1 PURPOSE OF DATA COLLECTION/GENERATION OR RE-USE AND ITS RELATION TO THE OBJECTIVES OF THE PROJECT	9
2.2 TYPES AND FORMATS OF COLLECTED/GENERATED DATA	11
2.3 ORIGIN OF DATA AND REUSE OF PRE-EXISTING DATA	16
2.4 EXPECTED SIZE OF DATA	19
2.5 DATA UTILITY	22
3. FAIR DATA	23
3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA	23
3.2 MAKING DATA ACCESSIBLE	26
3.3 MAKING DATA INTEROPERABLE	36
3.4 INCREASE DATA REUSE	37
4. OTHER RESEARCH OUTPUTS	43
5. ALLOCATION OF RESOURCES	47
5.1 ESTIMATED COSTS FOR MAKING DATA FAIR	47
5.2 DATA MANAGEMENT RESPONSIBILITIES	48
6. DATA SECURITY	52
7. ETHICS AND OTHER ISSUES	54
8. CONCLUSIONS AND WAY FORWARD	55
ANNEXES	56
ANNEX I – PRIVACY POLICY	56
ANNEX II – INFORMED CONSENT FORM	62
ANNEX III – DATA SUBJECT REQUEST FORM	64
ANNEX IV- MAIN CHANGES IN THE DATA COLLECTED/GENERATED SINCE M4	67

List of Figures

Figure 1: Typical DOI created by Zenodo	27
Figure 2: CC BY-SA 4.0.....	38
Figure 3: CC BY 4.0	38
Figure 4: CC BY-ND 4.0.....	38
Figure 5: CC BY-NC 4.0	38
Figure 6: CC BY-NC-ND 4.0	39

List of Tables

Table 1: Terms and Definitions	4
Table 2: MANOLO Partners.....	6
Table 3: Open-source software and datasets	17
Table 4: Expected size of data.....	19
Table 5: Data Utility.....	22
Table 6: Good practices for data anonymisation.....	28
Table 7: Data Availability.....	28
Table 8: Data Accessibility	32
Table 9: Dublin Core Metadata standard vocabulary.....	36
Table 10: Expected time that the data will be made open	39
Table 11: Other digital or physical research outputs	43
Table 12: Estimated costs for making data FAIR	47
Table 13: Data management responsibilities of MANOLO' partners per data collected/generated under each WP.....	50

List of Terms and Definitions

Table 1: Terms and Definitions

Acronym	Definition	Acronym	Definition
AI	Artificial Intelligence	ML	Machine Learning
CC BY-SA 4.0	Creative Commons Attribution-Share Alike 4.0	OAI	Open Archives Initiative
CC BY 4.0	Creative Commons Attribution 4.0 International	OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
CC BY-ND 4.0	Creative Commons Attribution-No Derivatives 4.0 International	PID	Persistent Identifier
CC BY-NC 4.0	Creative Commons Attribution-Non-Commercial 4.0 International	PC	Project Coordinator
CC BY-NC-ND 4.0	Creative Commons Attribution-Non-commercial-No Derivatives 4.0 International	PMO	Project Management Office
CESSDA	Consortium of European Social Science Data Archives	PO	Project Officer
COTS	Commercial-Off-The-Shelf	QA	Quality Assurance
CPU	Central Processing Unit	QC	Quality Control
DCMI	Dublin Core Metadata Initiative	ROS	Robot Operating System
DNN	Deep Neural Network	SMA	Social Media Accounts
EEA	European Economic Area	TL	Task Leader
FAIR	Findable, Accessible, Interoperable and Re-usable	URL	Uniform Resource Locator
GDPR	General Data Protection Regulation	WP	Work Package
HW	Hardware	WPL	Work Package Leader
KPI	Key Performance Indicator	WTL	Work Task Leader
DMP	Data Management Plan		

Executive Summary

This document constitutes the interim version of the **Data Management Plan (DMP)** and has been elaborated as a deliverable (D8.3) in the framework of the MANOLO project. MANOLO will deliver a complete stack of trustworthy algorithms and tools to help AI systems achieve better efficiency and seamless optimization in their operations, resources, and data required to train, deploy, and run high-quality and lighter AI models in centralised and cloud-edge distributed environments. Under this light, MANOLO entails several activities within its framework that involve the collection, production, and/or processing of data, intending to generate meaningful insights that will feed into the project and push the state of the art by innovating a suite of complementary algorithms for training, comprehension, compression, and optimization of models. This involves advancing research in the areas of model compression, meta-learning (few-shot learning), domain adaptation, frugal neural network search and growth, and neuromorphic models.

In this context, the interim version of the project's DMP reviews and revises, when necessary, the overall methodological principles pertaining to the management of the data that will be collected, generated and/or re-used in the framework of MANOLO, safeguarding sound and ethical data management along the entire duration of the project. Moreover, it provides an updated overview of MANOLO's data, as identified in this interim stage of the project, along with information on the methodology about their management as well as making them Findable, Accessible, Interoperable and Re-usable (FAIR). The methodology of MANOLO for data management builds on know-how, tools, and templates that were developed internally by Q-PLAN as well as on good practices and templates from the literature (such as the Horizon Europe Data Management Plan Template). As in previous EU-funded projects, tailored modifications to the methodology were implemented for MANOLO as well, to comply with the GA conditions and the particularities of the project. Along these lines, this deliverable presents the adjusted methodology as it was further developed and applied in the context of MANOLO, as well as presents the results from its application during the project.

The interim version of the DMP is the second of the three versions of MANOLO's Data Management Plan to be produced in the course of the project and will serve as living document (D8.2 Data Management Plan v.1 delivered in M4 has been updated to D8.3 Data Management Plan v.2 in M18 and will be ultimately fixed as D8.4 Data Management Plan v.3 in M36). Along these lines, the initial DMP has been updated and will be further elaborated during the project to reflect an accurate, up-to-date, and ultimately comprehensive plan for managing the data that will be collected, generated and/or re-used by the project across their entire life cycle, both during and after the completion of MANOLO.

The methodology of MANOLO for data management builds on know-how, tools, and templates that were developed internally by Q-PLAN as well as on good practices and templates from the literature (such as the Horizon Europe Data Management Plan Template). As in previous EU-funded projects, tailored modifications to the methodology were implemented for MANOLO as well, to comply with the GA conditions and the particularities of the project. Along these lines, this deliverable presents the adjusted methodology as it was further developed and applied in the context of MANOLO, as well as presents the results from its application during the project.

1. Introduction

1.1 Objectives

The current document represents the interim version of MANOLO's Data Management Plan (DMP), which received funding from the European Union's Framework Programme for Research and Innovation Horizon Europe under Grant Agreement No 101135782.

MANOLO aims to provide a robust **set of trustworthy algorithms and tools to optimize Artificial Intelligence (AI) systems**, focusing on the efficient development, deployment, and operation of high-quality, lightweight models in both centralized and cloud-edge environments. To achieve this, MANOLO will develop a versatile toolset with algorithms focusing on training, understanding, compressing, and optimizing machine learning models. Thus, the methodology that MANOLO is going to follow involves advancements in model compression, meta-learning, domain adaptation, frugal neural network search, and growth, along with the development of dynamic algorithms for policy-compliant AI task allocation in the cloud-edge continuum, ensuring widespread deployment with trustworthiness.

To this end, the **consortium** of MANOLO brings together a complementary and interdisciplinary group of **18 partners across 8 different countries** within the EU, as presented in the table that follows.

Table 2: MANOLO Partners

Partner Role*	Partner No	Partner Name	Partner Short Name	Country
CO	1	UNIVERSITY COLLEGE DUBLIN, NATIONAL UNIVERSITY OF IRELAND, DUBLIN - CeADAR Ireland's Centre for AI	NUI DUCD-CeADAR	Ireland
BEN	2	UNIVERSITAT POLITECNICA DE CATALUNYA	UPC	Spain
BEN	3	ATOS IT SOLUTIONS AND SERVICES IBERIA SL	ATOS IT	Spain
AE	3.1	EVIDEN TECHNOLOGIES SRL	EVIDEN RO	Romania
BEN	4	TECHNISCHE UNIVERSITAET BRAUNSCHWEIG	TUBS	Germany
BEN	5	NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS"	NCSR "D"	Greece
BEN	6	FOUR DOT INFINITY INFORMATION AND TELECOMMUNICATIONS SOLUTIONS PRIVATE COMPANY	FDI	Greece
BEN	7	FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV	Fraunhofer	Germany
BEN	8	INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE	INRIA	France
AE	8.1	UNIVERSITE PARIS-SACLAY	UPSaclay	France
BEN	9	YRKESHOGSKOLAN ARCADA AB	ARCADA	Finland
BEN	10	KATHOLIEKE UNIVERSITEIT LEUVEN	KU Leuven	Belgium
BEN	11	LAUREA-AMMATTIKORKEAKOULU OY	LAU	Finland
BEN	12	PAL ROBOTICS SL	PAL ROBOTICS	Spain
BEN	13	BIT & BRAIN TECHNOLOGIES SL	BIT&BRAIN	Spain

Partner Role*	Partner No	Partner Name	Partner Short Name	Country
BEN	14	ARX NET AE YPIRESIES KAI EPICHIRISIS DIADIKTYOU ANONIMI ETAIRIA	ARX.NET S.A.	Greece
BEN	15	Q-PLAN INTERNATIONAL ADVISORS PC	Q-PLAN	Greece
BEN	16	EIT DIGITAL	EIT DIGITAL	Belgium

* CO = Coordinator, BEN = Beneficiary, AE = Affiliated Entity

All partners of MANOLO's consortium adhere to sound data management principles to ensure that the meaningful data collected, generated, and/or reused throughout the project are well-managed, archived, and preserved, in line with the structure and guidelines of the Horizon Europe Data Management Plan Template¹.

Hence, this interim version of the DMP aims to achieve the following objectives:

- Describe the data management lifecycle for the data to be collected, generated, and/or reused in the framework of MANOLO, serving as the key element of good data management.
- Outline, review and revise the methodology employed to safeguard the sound management of the data collected, and/or generated, as well as to make them Findable, Accessible, Interoperable, and Reusable (FAIR).
- Provide updated information on the data that will be collected, generated, and/or reused and how it will be handled during and after the end of the project, along with the standards applied to this end.
- Describe details on how the data will be made openly accessible and searchable to interested stakeholders, as well as its curation and preservation.
- Address the management of any research outputs other than data in line with FAIR principles.
- Present updated information on the resources to be allocated to make data FAIR, clearly identifying responsibilities of data management, while addressing data security and ethical aspects.

With the above in mind, this interim version of **the DMP is structured in 8 distinct chapters**, as follows:

- **Chapter 1** provides introductory information about the DMP, the context in which it has been elaborated, as well as its objectives and structure.
- **Chapter 2** presents a summary of the data to be collected/generated, or reused during the activities of MANOLO, including its purpose as well as its types and formats. Additionally, it outlines its origin, expected volume, and the stakeholders that may find it useful.
- **Chapter 3** describes the methodology that is applied in MANOLO to safeguard the effective management of data across its entire lifecycle, making it FAIR.
- **Chapter 4** presents the management of other research outputs that may be generated or reused throughout MANOLO and provides sufficient details on making them FAIR.
- **Chapter 5** estimates the resources required for making the project's data FAIR, while also identifying data management responsibilities.
- **Chapter 6** outlines the data security strategy applied within the context of MANOLO, along with the respective secure storage solutions employed.
- **Chapter 7** addresses ethical aspects as well as other relevant issues about the data collected/generated, or reused during the implementation of the project.
- **Chapter 8** concludes with the next steps foreseen in the framework of the project concerning its data management plan.

¹ https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

Annexed in the document are (i) the project's Privacy Policy (Annex I), the templates for the (ii) Informed Consent Form (Annex II), (iii) the Data Subject Request Form (Annex III), and (iv) the main changes in the data collected/generated or re-used since their original definition.

Note that **the DMP is not a fixed document**. It evolves during the lifespan of the project and will be further elaborated and **updated at least once more throughout the duration of MANOLO (i.e.D8.4 in M36)**. Additional ad hoc updates may also be realised (if necessary), to include new data, better detail, and/or reflect changes in the methodology or other aspects relevant to their management (such as costs for making data FAIR, size of data, etc.), changes in consortium policies and plans or other potential external factors. **Q-PLAN is responsible for the elaboration of the DMP and, with the support of all partners, will update and enrich it when required.**

2. Data Summary

MANOLO collects/generates or reuses meaningful non-sensitive data that do not fall into any special categories² of personal data as those described within the General Data Protection Regulation³ (GDPR). This data may be quantitative, qualitative, or a blend of those in nature and will be analysed from a range of methodological perspectives to produce insights that successfully feed MANOLO's activities, enable us to deliver evidence-based results, and ultimately achieve the objectives of the project. With that in mind, the second chapter of the Data Management Plan (DMP) starts by explaining the purpose for which this data will be collected/generated and how it relates to MANOLO. It proceeds by describing the different types and formats of this data as well as their origin and expected volume, before concluding with an overview of potential stakeholders for whom it may prove useful for re-use.

2.1 Purpose of data collection/generation or re-use and its relation to the objectives of the project

To successfully meet its objectives and ensure the production of evidence-based results, MANOLO entails several activities during which data are collected/generated or reused. The purpose for which this data are collected/generated, or reused, is interrelated with the objective of the activity during which it is produced.

In particular, these activities, along with their objectives, in the framework of MANOLO are as follows:

1. **Identification of insights about the needs, challenges, available or forthcoming technological solutions, key actors, and openly available resources.** With mapping (desk research) relevant online sources and databases regarding AI data and the cloud-edge continuum, interviewing key public and private stakeholders within the consortium but also with external entities derived from the desk research and conducting an online survey that will target end-users, technology providers, data experts, cloud-edge continuum experts, and decision-makers.
2. **The development of a benchmarking framework entails the establishment of a comprehensive process for co-design, encompassing both training and inference models across the entire edge-cloud continuum.** This involves defining specific functional and non-functional requirements, including considerations for social and privacy aspects. The aim is to facilitate the seamless implementation, experimentation, deployment, evaluation, and validation of the envisioned solution.
3. **Automatically annotating data concerning quality or possible unreliability.** Methods for anomaly detection and noise detection will be implemented to identify biased, noisy, inconsistent, or low-quality data.
4. **Optimising data and models for AI by distilling key information, compressing data, extracting relevant features, synthesizing additional data points, transferring knowledge between models, and pruning unnecessary parameters.** This approach aims to produce high-quality data and extreme model compression without sacrificing performance, facilitating the development of novel algorithms and efficiently deploying AI solutions in resource-constrained environments.

² Special categories of personal data according to Regulation (EU) 2016/679 of the European Parliament (General Data Protection Regulation) include personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

³ Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

5. **Combining techniques like knowledge distillation, quantization, and weight pruning for Compressing models.** Knowledge distillation transfers knowledge from a large model (teacher) to a smaller one (student), reducing size and complexity while maintaining performance. Quantization reduces weight representations, saving memory and computation. Weight pruning removes unnecessary connections, further reducing model size. These techniques extend to few-shot learning, using compressed data for further quantization and pruning, ultimately achieving extreme model compression while maintaining performance.
6. **Domain adaptive Few-Shot Continuous Learning by developing algorithms for meta-learning and model compression,** leveraging techniques from T3.1. The goal is to optimize models for multiple domains with efficient few-shot continuous learning.
7. **Computational resources and AI assets across the distributed heterogeneous cloud-edge system will be mapped, abstracted, and categorized.** This involves collecting information about resource capabilities, networking topology, and AI functions. And designing/ developing a policy compliance manager that ensures AI/ML functions and models conform to predefined and configurable criteria/policies.
8. **Identifying expressivity bottlenecks and guiding neural network growth.** Training data are reused to estimate expressivity bottlenecks, detect spurious correlations, and evaluate network expansions. Metrics about model performance and computational efficiency are collected during the iterative architecture growth process.
9. **Training and optimizing spiking neural networks (SNNs) using knowledge distillation.** Existing trained DNN models and limited training datasets are reused to distill knowledge into hardware-specific SNN student models. Simulation results and performance metrics (e.g., accuracy, energy consumption, latency) are collected to evaluate and optimise the student networks.
10. **Designing and developing optimized mechanisms for allocating AI/ML models and functions across the edge-cloud continuum.** It considers the specific requirements of the continuum infrastructure, such as heterogeneous and distributed computational resources with different capabilities and limitations, as well as policy validation and federated training functionality. The goal is to model parameters characterizing resources and associated policies to optimize the allocation of ML models and functions.
11. **Extend the KOBE Open Benchmarking Engine to apply ML workloads to systems with heterogeneous hardware components and dynamic network connectivity.** Orchestrate experiments that benchmark Neural Network models on Deep Neural Network (DNN) and Spiking Neural Network (SNN) neuromorphic accelerators, as well as Commercial-Off-The-Shelf Hardware (COTS-HW), to validate benchmarking results against stress tests.
12. **Implement a module for analyzing energy efficiency and performance of AI/ML models.** Evaluate energy consumption, accuracy, precision, specificity, speed, and performance for integration into the MANOLO benchmark framework by providing key performance indicators (KPIs) such as latency and energy consumption after running models on neuromorphic accelerators and COTS-HW.
13. **Develop methods and tools for human oversight in monitoring and validating AI systems** within the cloud-edge continuum. Use the Z-Inspection® process for Trustworthy AI assessments, ensuring ongoing compliance with trustworthiness criteria.
14. **Deploying MANOLO's framework in four use cases,** integrating technology, preparing testbeds, and conducting final evaluations under realistic conditions.
15. **Evaluating and validating the MANOLO solution** based on data generated from the use case deployment. Addressing technical and non-technical aspects, including learning accuracy, MPC performance, energy consumption, ethics, safety, and trustworthiness.
16. **Monitoring and assessment of the dissemination, communication, stakeholder engagement, and clustering activities** of MANOLO, to measure their results and impact, fine-tune our strategy in this respect as well as fulfil the project's reporting requirements towards the Commission.

17. **Business modelling and planning exercise which includes market analysis** (market size, market structure, prospects, and growth rate, etc.), **competition analysis** (number and profile of competitors, potential gap, etc.), **as well as financial planning** (cash flow analysis, costs, and result forecast, etc.). This exercise will fuel the elaboration of the MANOLO Business Plan, which aims to ensure the effective and viable exploitation of the MANOLO results after the end of the project.

18. **Setting Up and Maintaining the Ethics & Industrial Advisory Board (EIAB).**

The following section provides further details on the different types and formats of data collected/generated, or reused during the project's activities.

2.2 Types and formats of collected/generated data

MANOLO is set to collect/generate data of various structures and formats. Along these lines, the data definition process used for this DMP is based on the source and the physical format of the data⁴. In particular, we define two main aspects: (i) the process under which the underlying data are created/captured, which includes electronic text documents, spreadsheets, questionnaires, images, videos, code, and transcripts, among others, and (ii) the storage format of quantitative and qualitative data. Examples of this aspect include easily accessible formats, such as post scripts (e.g., .pdf, .xps, etc.), machine-readable formats (.xml, .html, .bag, etc.), spreadsheets (e.g., .xlsx, .csv, etc.), text documents (e.g., .docx, .rtf, etc.), images (e.g., .jpeg, .png, etc.), videos (e.g., .mp4, .avi, etc.), code (e.g., .cpp, .py, .lib, .rospack, etc.), compressed formats (e.g., .rar, .zip, etc.) or any other format required by the objectives and methodology of the activity within the framework of which is produced.

Under this framework, special attention will be paid to using open **formats**⁵ (such as .csv, .pdf, .zip, etc.) and/or **machine-readable formats**⁶ (such as .XML, .JSON, .rdf, .html, .bag, .py, .cpp, etc.) when possible, to enhance the **interoperability and re-use of data**. In doing so, we are providing data that are **easily readable** and **freely usable in any software program** employed by third parties interested in utilizing the data.

The type and formats of the data collected/generated in the context of MANOLO can be divided into 3 categories, namely:

- data collected/generated by **direct input methods**;
- data collected/generated through **the use of the MANOLO framework, and its development**; and
- data collected/generated **from dissemination, communication, and stakeholder engagement activities**.

These categories are described in more detail in the following subsections.

2.2.1 Data collected/generated through direct input methods

Direct input methods, under the scope of MANOLO, involve methodologies for collecting data through desk research and interactions between consortium partners and external stakeholders, with the latter providing data to the former. Along these lines, external stakeholders undertake the role of a data subject, which is a

⁴ Jakobsson, U., Braukmann, R., Lundgren M., Expert Tour Guide on Data Management. Retrieved from <https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/1.-Plan>.

⁵ According to the [Open Data Handbook](#): "An open format is a file format with no restrictions, monetary or otherwise, placed upon its use and can be fully processed with at least one free/open-source software tool and it is not encumbered by any copyrights, patents, trademarks or other restrictions so that anyone may use it".

⁶ According to the [Open Data Handbook](#): "Machine readable formats are file formats that can be automatically read and processed by a computer. Machine-readable data must be structured data".

natural person whose personal data are being processed⁷. In particular, the identification and selection of suitable data subjects are based on purposeful sampling according to which, external stakeholders are identified and selected by consortium partners based on their relevance with the MANOLO framework (e.g., AI, cloud-edge environments, etc.) and the objectives of the respective activity for which data are collected. In this context, quantitative and qualitative data⁸ will be collected/generated during MANOLO:

- **Quantitative data** are numerical and acquired through counting or measuring. Examples of quantitative data are the yearly turnovers of a business, the hourly compensation of a worker, the number of SMEs in Europe, etc. This data may be represented by ordinal, interval, or ratio scales and lend themselves to statistical manipulation.
- **Qualitative data**, sometimes referred to as categorical data, are data that can be arranged into categories based on physical traits, gender, colours, or anything that does not have a number associated with it. Moreover, written documents, interviews, and various forms of in-field observation are all sources of qualitative data. Examples of qualitative data are the preferences for learning, skillsets, country of origin, etc.

Additional details concerning the different types and formats of data that will be collected through direct input methods under the frame of MANOLO are provided below.

Identification of insights about the needs, challenges, available or forthcoming technological solutions, key actors, and openly available resources

The data that were collected/generated here fall under Task 1.1 (Q-PLAN). The data were quantitative and qualitative data on mapping (desk research) relevant online sources and databases regarding AI data and the cloud-edge continuum, interviewing key public and private stakeholders within the consortium, but also with external entities derived from the desk research and by conducting an online survey that targeted end-users, technology providers, data experts, cloud-edge continuum experts, and decision-makers. The formats are .docx, .xlsx, .csv and .pdf.

Business model design and validation

Within the frame of Task 7.3 (Q-PLAN), business models will be designed to fuel the development of the MANOLO Business Plan, which aims to ensure the effective and viable exploitation of the MANOLO results after the end of the project. These business models will be validated through interviews with key MANOLO partners and/or an online survey with external relevant stakeholders. The data to be collected will be both qualitative and quantitative, and the expected format will be .xlsx and .docx.

Setting Up and Maintaining the Ethics & Industrial Advisory Board (EIAB)

KU Leuven in Task 8.2 is responsible for setting up an Ethics & Industrial Advisory Board. The EIAB provides the Consortium with valuable feedback to better align our outcomes with the needs of users and stakeholders and oversees any ethics issues that may arise during the project. Each partner identified and suggested suitable experts (3 - 5 per partner) to join the EIAB. The expected formats are .xlsx and .docx.

⁷ Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>.

⁸ Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches. Boston: Pearson.

2.2.2 Data collected/generated by the development and use of the MANOLO framework

This section focuses on the data collected and its formats. Additional documentation, accompanying each data set (such as a manifest listing each data in the dataset, or metadata to facilitate its usage can be found in Section 3.

Development of a benchmarking framework entails the establishment of a comprehensive process for co-design, encompassing both training and inference models across the entire edge-cloud continuum

This process in Task 1.3 (ATOS IT) includes defining functional and non-functional requirements such as memory, connection throughput, location of input and output data, and energy efficiency. It aims to enable seamless implementation, experimentation, deployment, evaluation, and validation of the solution. The framework supports structured and multimodal time-series data and uses a text-based format, like line-oriented/JSON, for input and output data.

Automatically annotating data with respect to quality or possible unreliability

The activity in Task 2.2 (NCSR "D") aims to create a framework that provides high-quality data and metadata for training AI models efficiently, especially when original data is limited or unsuitable for the desired architecture. Key objectives include data distillation, data compression and hashing, feature extraction and synthesis, and model inversion to synthesize data from labels. These methods are being tested on collected datasets, integrating with the Data Inspection & Generation component from another task. The framework's outcomes will support novel algorithm development by providing datasets in formats such as pickle, .csv, .jpg, .png, and .txt for various types of data, including image data, signal data, structured and multimodal time series.

Data Distillation & Synthetic Feature Representation

NUIDUCD - CeADAR in Task 2.3 will generate new datasets either by feature extraction from existing data, synthetic data generation, or data distillation techniques. Supporting metadata will be provided. The expected formats are Pickle, .csv, .jpg, .png, and .txt.

Model Pruning, Quantization, and Knowledge Distillation

Modelling activities in Task 3.1 (NUIDUCD - CeADAR) will generate data as model results and statistics. This data will be used as WP3 for decision-making and to keep improving the models produced. The expected formats are .txt and .csv.

Domain Adaptive Few-Shot Continuous Learning

Modelling activities in Task 3.2 (NUIDUCD - CeADAR) will generate data as model results and statistics continuously over time. Thus, it will be a structured table with qualitative and quantitative data regarding compression techniques and resulting metrics for different models. The expected formats are .txt and .csv.

Computational resources and AI assets across the distributed heterogeneous cloud-edge system will be mapped, abstracted, and categorized.

This activity in Task 4.2 (TUBS) focuses on mapping, abstracting, and categorizing computational resources and AI assets across a distributed heterogeneous cloud-edge system infrastructure. It considers various aspects of resource capabilities, such as processing, storage, energy, capacity, and networking topology, as well as associated AI functions and assets in the cloud-edge environment. The information gathered will be used to define specific mapping parameters that, along with a policy validator from another task, will support the development of dynamic model allocation mechanisms across the continuum. This includes characterizing

nodes based on architecture, CPU, memory, storage, and handling AI applications using formats like text, code, JSON, and YAML.

Use case AI applications and their requirements

The use case AI applications in Task 4.2 (TUBS) involve describing AI application requirements, such as the use of PODs, Kubernetes, containers, dockers, and the location of input and output data. The resource topology includes understanding the available resources and their characteristics. AI applications are represented using text, code, JSON, and YAML formats.

AI/ML Models and functions allocation on the cloud-edge continuum

Designing and developing optimized mechanisms in Task 4.3 (ATOS IT) for allocating various AI/ML models and functions (training/inference) across the cloud-edge continuum. The information about nodes in the cloud-edge after applying optimization algorithms will be provided. The expected formats are .xlsx, .docx, and .JSON.

Extend the KOBE Open Benchmarking Engine to apply ML workloads to systems with heterogeneous hardware components and dynamic network connectivity

The activity in Task 5.1 (NCSR "D") extends the KOBE Open Benchmarking Engine to apply ML workloads to systems with heterogeneous hardware components and dynamic network connectivity. This includes handling different memory types, varying connection throughputs, and the locations of input and output data. The goal is to optimize the benchmarking process for diverse system architectures and network configurations and orchestrate experiments, allowing for efficient ML applications across different resource types like LaTeX or text.

Implement a module for analyzing energy efficiency and performance of AI/ML models

The module in Task 5.2 (ATOS IT) calculates key performance indicators (KPIs) such as latency, energy consumption, accuracy, precision, specificity, and speed. It assesses electric power and energy consumption metrics on bare metal hosts, providing structured time series and text-based data in line-oriented or JSON format. The module's output integrates with the MANOLO benchmark framework to offer insights into model performance, enabling comparison across different hardware configurations and models for improved decision-making in AI/ML development.

Develop methods and tools for human oversight in monitoring and validating AI systems within the cloud-edge continuum

The activity in Task 5.5 (ARCADA) involves developing methods and tools for human oversight in monitoring and validating AI systems within the cloud-edge continuum. This includes creating templates and self-assessment tools for evaluating AI systems to ensure ethical and reliable performance. These resources will be available in text documents in formats such as .docx and .pdf.

Interaction with humans for assisted living

The activity in Task 6.2 (BIT&BRAIN) focuses on enabling interaction with humans for assisted living scenarios. This includes conversations and recordings stored as .mp3 files, as well as robot state information such as joint positions and status details from ROS 2 topics. Also involves managing 3D environment information through point cloud data (.pcd) and 2D environment information through RGB images (.png/.jpeg). The goal is to create a seamless integration of human-robot interaction for improved assisted-living experiences.

Mobile computer vision

Task 6.2 (BIT&BRAIN) involves mobile computer vision applications, focusing on image classification, face detection, and manual tagging (ARX.NET). It includes processing photographs in .jpeg format to enhance

mobile vision capabilities, supporting tasks such as identifying and categorizing images, recognizing faces, and adding manual tags for improved data organization and analysis.

Auditory stimulation for memory consolidation

In the context of Use Case 3 (BIT&BRAIN), in Task 6.2 (BIT&BRAIN), we will demonstrate MANOLO's technology for the implementation of at-home treatments based on physiological signals (e.g. electrophysiological data). We will use already recorded and anonymized datasets for the development, fine-tuning, and preparation of the use case. The data will consist mainly of sleep recordings and the corresponding sleep stages annotated by experts in .csv format.

Evaluation and Validation of MANOLO Solution

Data generated in Task 6.3 (LAU) from deploying use cases in the health, telecommunications, and manufacturing sectors will be the primary basis for evaluation and validation activity. Additionally, data generated by other work packages (WP2-WP5) will also be considered during this activity. The expected formats are .xlsx and other possible formats.

2.2.3 Data collected/generated from dissemination, communication, and stakeholder engagement activities

The data deriving from monitoring and assessment of the dissemination and communication results of the project and by the stakeholder engagement to measure the impact of the relevant activities, fall under Task 7.1 (Q-PLAN) and include (i). Website and social media accounts (i.e., YouTube, LinkedIn, X (Twitter), at this point in the project) analytics; (ii). Data collected from project events; (iii) Data collected from dissemination and communication actions (e.g., participation in external events, participation in project workshops, etc.); and (iii). Newsletter subscriptions. The data are identified through online analytics, utilising Google Analytics, SMAs, partners' reporting, Mailchimp platform, etc. Q-PLAN is responsible for sending the necessary templates (.docx, .xlsx) to all partners, alongside guidelines on how to fill them in, as well as for collecting input on an ad-hoc basis (i.e., each time a dissemination or stakeholder engagement action is performed). Q-PLAN is also responsible for preparing the necessary reports to evaluate the overall progress of dissemination and communication activities (measuring outcomes against pre-set KPIs) throughout the lifespan of the project. The storage format of the data includes .csv, .docx, .xlsx, .pdf, .ppt, .jpeg, and .png files.

Website and social media analytics

Website analytics: These data are being collected/generated through periodic monitoring of the project's website statistics to measure and assess the performance and results of the project's website activity in terms of dissemination and communication. With that in mind, the data are mostly quantitative in nature, addressing key metrics (e.g., number of visitors, average page views per session, session duration, average time on page, country of visitors, occupation of visitors, etc.). All in all, the data are stored in spreadsheets (.xlsx or .csv) while at the same time, the analysis of the results is stored in a standard Word document (.docx).

Social media statistics: These data are collected/generated through periodic monitoring of the project's social media statistics (i.e., YouTube, Twitter, and LinkedIn) to measure and assess the performance and results of the project's social media activity in terms of dissemination and communication. With that in mind, the data are mostly quantitative in nature, addressing the metrics reached on each channel (e.g., number of followers, tweet impressions on Twitter, friends, likes on LinkedIn, number of people reached through posts, etc.). All in all, the data are stored in a Microsoft Excel file (.xlsx) while at the same time, the analysis of the results is stored in a standard Word document (.docx).

Data collected from project events

These data are being collected during the implementation of the project through: (i) the different events (e.g. conferences, interviews, physical and virtual events, etc.) organised by MANOLO (either alone or jointly with other projects or initiatives) consisting of the participant's lists that enclose demographic information about the participants; and (ii) the participation of MANOLO partners in relevant third party events in order to reach out and engage stakeholders, thus collecting general information about the events attended and their outreach.

Along these lines, these data are being collected to keep track of the results of activities in events for stakeholder engagement and provide the opportunity for project partners to report on these activities. Moreover, these data are updated every time a partner attends an event or a partner organises an event. Finally, the data are both quantitative and qualitative in nature and are stored in a standard spreadsheet (.xlsx).

Data collected from dissemination and communication actions

These data are being collected through the periodic monitoring of the project's miscellaneous dissemination activities such as publications in relevant journals, posts in blogs, etc. The data consist of a spreadsheet (i.e., .xlsx or .csv) designed to keep track of any kind of communication and dissemination activity, including, but not limited to, press releases, social media posts, website articles, interviews, events (conferences, meetings, workshops, etc.), other publications, e-mails, presentations, informal discussions, seminars, etc. The purpose of collecting these data is to assess the outreach and efficiency of the dissemination activities during the implementation of the project. For this purpose, a template is being shared with all partners to recommend activities to be performed and log the activities they perform. The template is also provided online so that the partners can directly update their input. Finally, all the data are integrated into a single spreadsheet (.xlsx or .csv).

Newsletter subscriptions

A subscription form hosted on the project's website aids the collection of these data, in which any interested stakeholder can freely provide their contact details in a dedicated sign-up form to receive the most up-to-date news and outcomes of the project. A newsletter is being sent to subscribers once per 6 months. With that in mind, these data are being collected so that interested stakeholders can be informed about MANOLO's progress and upcoming events. The data comprise a list of subscribers along with their personal information such as: (i) email address, (ii) first and last name, (iii) country, (iv) type of organisation, and (v) region. A copy of this contact list is stored on MailChimp's server (<http://mailchimp.com>), which is used for email campaigns and newsletter distribution. All personal information included in this contact list is used and protected according to MailChimp's Privacy Policy. Subscribers can find details about how their personal information is managed on the respective privacy policy section of the MANOLO website.

2.3 Origin of data and reuse of pre-existing data

In the context of MANOLO, **new data** are being and will be collected/generated by partners as well as external stakeholders participating in the activities of the project. With that in mind, and aside from consortium partners, **external groups of stakeholders from which new data originate include:**

1. AI Researchers and Experts:

- Experts in various AI disciplines, including model compression, meta-learning, domain adaptation, and frugal neural network search.
- Researchers specializing in dynamic algorithms for policy-compliant AI task allocation in the cloud-edge continuum.

2. Industry and Technology Developers:

- Professionals and organizations involved in the development of AI technologies in both centralised cloud and edge environments, especially those focusing on efficient AI model development, deployment, and operation.
- Companies with expertise in AI toolsets and algorithms for optimizing machine learning models.

3. Academic Institutions and Research Organizations:

- Researchers and experts affiliated with academic institutions and non-university public research organizations specializing in AI, machine learning, and related fields.

4. Government and Regulatory Bodies:

- Policy makers and consultants at regional, national, and international levels are involved in shaping regulations related to AI ethics, model deployment, and trustworthy AI systems.

5. Technology Innovators and Startups:

- Innovative startups and businesses focusing on AI solutions, particularly those interested in adopting and implementing lightweight, high-quality models in their applications.

6. Collaborative Initiatives and Networks:

- Participation in relevant collaborative initiatives, networks, and projects related to AI, machine learning, and technology optimization.
- Engagement with initiatives that align with MANOLO's goals and contribute to the widespread deployment of trustworthy AI.

7. End-Users and Application Developers:

- Individuals and organizations involved in deploying AI models for various applications, spanning from centralized systems to edge devices.
- Developers and practitioners working on applications that require efficient and optimized AI models.

Moreover, **pre-existing data** are being utilised within the context of MANOLO. In particular, outputs from EU-funded projects (e.g., AGIMUS, ICOS, UPCAST, mF2C, CyberSecPro, etc.), national projects, institutions, and other relevant initiatives (ADRA, EDIHs, BDVA, EurAI, etc.), to a large extent, provide a solid basis for MANOLO. The MANOLO consortium strives to make the most of and advance the work and results of these projects and initiatives. Moreover, relevant data are being harvested from selected trusted online sources regarding the state of play on agile production in European industries, with a focus on the consortium countries. Table 3 provides a comprehensive overview of the open-source data and software that are being utilized throughout the project's duration. Finally, consortium partners' internal knowledge, experience, and expertise from their participation in other projects and initiatives directly and indirectly support the implementation of activities throughout the project.

Table 3: Open-source software and datasets

Name	Software/Dataset	Link	Notes
Tensorflow	Software	https://www.tensorflow.org/	
Pytorch	Software	https://pytorch.org/	
Keras	Software	https://keras.io/	
Flower	Software	https://flower.readthedocs.io/en/latest/	
Dalex	Software	https://dalex.drwhy.ai/	xAI

Name	Software/ Dataset	Link	Notes
HolisticAI	Software	https://holisticai.readthedocs.io/en/latest/	xAI
LIME	Software	https://lime-ml.readthedocs.io/en/latest/	xAI
LIME- TimeSeries	Software	https://github.com/mdhabibi/LIME-for-Time-Series/	LIME specifically for time series domain
Grad-CAM	Software	https://github.com/jacobgil/pytorch-grad-cam	xAI
scikit-learn	Software	https://scikit-learn.org/stable/	
scipy	Software	https://scipy.org/	
pillow	Software	https://pillow.readthedocs.io/en/stable/	
psutil	Software	https://psutil.readthedocs.io/en/latest/	
matplotlib	Software	https://matplotlib.org/	
seaborn	Software	https://seaborn.pydata.org/	
numpy	Software	https://numpy.org/	
tqdm	Software	https://tqdm.github.io/	
pandas	Software	https://pandas.pydata.org/	
mne	Software	https://mne.tools/stable/index.html	
optuna	Software	https://optuna.org/	
BNCI Datasets	Dataset	http://bnci-horizon-2020.eu/database/data-sets	
MLPerf Tiny	Dataset	https://github.com/mlcommons/tiny	Use Case 1: Keyword Spotting → Speech Commands Dataset Use Case 2: Visual Wake Words → Visual Wake Words Dataset Use Case 3: Image Classification → Cifar10 Dataset Use Case 4: Anomaly Detection → ToyADMOS Dataset
Bearing Data Center	Dataset	https://engineering.case.edu/bearingdatacenter	Use Case: Ball bearing test data for normal and faulty bearing → Seeded Fault Test Data
DaLiAc – Daily Life Activities	Dataset	https://www.mad.tf.fau.de/research/activitynet/dali-ac-daily-life-activities/	Use Case: Classification of activities → DaLiAc Dataset
Caltech- 101	Dataset	https://www.kaggle.com/datasets/imbikramsaha/caltech-101/data	Approximation of Use Case 3 dataset (Images)
birds	Dataset	https://www.vision.caltech.edu/datasets/cub_200_2011/	Vision Data Set
cars	Dataset	https://docs.pytorch.org/vision/master/generated/torchvision.datasets.StanfordCars.html	Vision Data Set
flowers	Dataset	https://www.robots.ox.ac.uk/~vgg/data/flowers/102/	Vision Data Set
animals	Dataset	https://www.kaggle.com/datasets/alessiocorrado99/animals10	Vision Data Set
gtsrb	Dataset	https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign	Vision Data Set
CWRU	Dataset	https://www.kaggle.com/datasets/brjapon/cwrubearing-datasets	Time Series Data Set
MINST	Dataset	https://www.kaggle.com/datasets/hojjatk/mnist-dataset	Vision Data Set
F-MINST	Dataset	https://www.kaggle.com/datasets/hojjatk/mnist-dataset	Vision Data Set

Name	Software/ Dataset	Link	Notes
RadioML2 018	Dataset	https://www.kaggle.com/datasets/pinxau1000/radio-ml2018/data	RF Data Set

2.4 Expected size of data

MANOLO entails a series of activities aiming at setting the stage for and ultimately facilitating the demand-driven, evidence-based development, piloting, evaluation, validation, and fine-tuning of the MANOLO framework and value propositions. With that in mind, the table that follows presents the different activities implemented during the course of the project in which data are collected/generated, the types and formats of the data, as well as the expected size of the data.

Table 4: Expected size of data

No	Name of activity	Data	Type of data	Format of data	Expected size of data ⁹
1	Identification of insights about the needs, available or forthcoming technological solutions, key actors, and openly available resources	Desk research and interviews	Text document	.docx/ .pdf	10MB
2		Survey's results	Text documents, Spreadsheets	.docx, .xlsx, .csv, .pdf	20MB
3		Data from the State of the Art and regulatory resources	Text documents, Spreadsheets,	.docx, .xlsx, .csv, .pdf	100MB
4	Development of a benchmarking framework entails the establishment of a comprehensive process for co-design, encompassing both training and inference models across the entire edge-cloud continuum	Memory, connection throughput, location of input and output data, energy efficiency	Structured time-series, multimodal time-series.	text-based format is line-oriented/J SON	100MB
5	Data Distillation & Synthetic Feature Representation	Use cases, datasets, and their derivatives	Image data, signal data, structured time series, and multimodal time series.	Pickle, .csv, .jpg, .png, .txt	5GB
6		Synthetic/Distilled dataset	Structured data, image data, compressed data, and text data.	Pickle, .csv, .jpg, .png, .txt	5GB
7	Model Pruning, Quantization, and Knowledge Distillation	Model results and statistics	Structured data	.csv, .txt	~GBs. This will depend on the number of experiments performed.
8		LLM dataset and its generation results	Text data		
9		T-Hyperparameter search on	Structured data	.csv, .txt	~GBs. This will depend on the

⁹ The estimated expected size of the data is based on the adjusted size of data generated via similar activities of project partners in the past unless otherwise indicated.

No	Name of activity	Data	Type of data	Format of data	Expected size of data ⁹
	Domain Adaptive Few-Shot Continuous Learning	compression algorithms, with results and dataset + model metadata			number of experiments performed.
10		Curated texts for LLMs	Specific text to adapt LLM domain expertise		
11		Generated responses from LLMs	Generative texts from LLM trustworthiness experiments (reference only)		
12	Computational resources and AI assets across the distributed heterogeneous cloud-edge system will be mapped, abstracted, and categorized.	Resource topology: Nodes characteristics such as Architecture, CPU, Memory, Storage	AI applications: text, code	.json .yaml	~GBs. This will depend on the use case infrastructure and the AI applications to be executed in the use cases.
13	Use case AI applications and their requirements should be described	Use case AI applications description: (PODs, Kubernetes, containers, dockers, etc) and location of input and output data	AI applications: text, code	.json .yaml	~GBs. This will depend on the use case infrastructure and the AI applications to be executed in the use cases.
14	AI/ML Models and functions allocation on the cloud-edge continuum	Deployment information after optimization	Spreadsheets, code, text	.xlsx, .docx, .json	21.54 KB
15	Implement a module for analysing energy efficiency and performance of AI/ML models	Electric power and energy consumption metrics consumed on bare metal hosts	Structured time-series	text-based format is line-oriented / JSON	100MB
16	Develop methods and tools for human oversight in monitoring and validating AI systems within the cloud-edge continuum	Templates and Self-assessments	Text documents	.docx/ .pdf	100MB
17	Interaction with humans for assisted living	Conversations	Recordings	.mp3	~GBs. This will depend on the needs of the project
18		Robot state information	Joint positions, status info	ROS 2 topics	~MBs
19		3D information of the environment	RGDB images	.pcd	~GBs. This will depend on the

No	Name of activity	Data	Type of data	Format of data	Expected size of data ⁹
					needs of the project
20		2D information about the environment	RGB images	.png/.jpeg	~GBs. This will depend on the needs of the project
21		Model's performance in different nodes of the topology	Structured time-series	.json	~GBs. This will depend on the needs of the project
22	Mobile computer vision; classification	Image classification, photograph tagging, and visual cues for predictions	Mobile photo album class ontology, RGB images	Metadata: .json, data: .png/.jpg	~GBs. This will depend on the needs of the project
23	Mobile computer vision; Face detection	Face detection, face retrieval	RGB images	Metadata: .json, data: .png/.jpg	~GBs. This will depend on the needs of the project
24		Sleep recordings for sleep stage decoding	2-channel EEG signals during a full night of sleep. N (up to 100 nights, but it will depend on informed consent). IMU recordings for head movement. Annotation by experts. Learned a DeepNet model for decoding.	.csv	Each night will consist of 6-8 hours of sleep. 4bytes * 2 EEG channels * 256 (sampling rate)* 3600 (per hour) * 7 (number of hours) + IMU
25	Auditory stimulation for memory consolidation	Sleep recordings during real-time closed-loop auditory stimulation	2-channel EEG signals during a full night of sleep. N to be defined (<10). Additional signals and events: • Sleep stages are automatically scored by the DeepNet. • Signals provided by a frequency detection algorithm, acting as an additional go/stop indicator for automatic stimulation in addition to the sleep stage. • Time points of detections and stimulations performed by the stimulation algorithm, in phase with the targeted events of brain wave activity.	.csv	Each night will consist of 6-8 hours of sleep. 4bytes * 2 EEG channels * 256 (sampling rate)* 3600 (per hour) * 7 (number of hours) + Additional signals and events
26	Evaluation and Validation of MANOLO Solution	Use cases, data evaluation, and validation	End-user-generated Machine-generated (where applicable) Data generated from other work packages (technical and non-technical data)	.xlsx (and other possible formats)	20 MB (plus or minus)
27	Monitoring and assessment of the	Website and social media analytics	User-generated/ machine-generated	.xlsx	20 MB

No	Name of activity	Data	Type of data	Format of data	Expected size of data ⁹
28	dissemination, communication, and stakeholder engagement activities	Data collected from project events	Spreadsheets, Event Reports (incl. minutes, participants lists), Photos, Video recordings, Event agendas, Event promotional material	.xlsx, .docx, .pdf, .jpg, .png, .pptx, .mov, .mp4	200 KB
29		Data collected from dissemination and communication actions	Spreadsheets, images	.xlsx, .pdf, jpeg, .png	200 KB
30		Newsletter subscriptions	Spreadsheets	.csv, .xlsx	100 KB
31	Business Modeling and planning	Market analysis, Competition analysis, as well as financial planning	Notes, spreadsheets	.docx, .xlsx	20 MB
32	Setting Up and Maintaining the Ethics & Industrial Advisory Board (EIAB)	Constituting the EIAB	Spreadsheet	.xlsx	100 KB
33		Management of the EIAB	Spreadsheet	.docx	1 MB

2.5 Data utility

The stakeholders that may find meaningful utility for the data to be collected/generated by the project (both within as well as outside of the MANOLO consortium), along with the benefits that could arise for them by utilizing these data, are concisely presented in the table that follows.

Table 5: Data Utility

Stakeholder group	Data utility
Academia & Researchers	Contribution to the advancement of AI research and the creation of more efficient and trustworthy AI systems.
Industries & Organisations	Reduced energy consumption and environmental footprint, as well as improved performance and competitiveness.
Government bodies and policymakers	The promotion of efficient and trustworthy AI systems will ensure that AI development aligns with European values and meets regulatory requirements.
Data providers	New opportunities for data providers and owners to contribute to AI development while ensuring data privacy and security

3. FAIR data

The guidelines on Data Management Plans available in the [Horizon Europe Data Management Plan Template](#) of the Commission emphasise the importance of making the data produced by projects funded under Horizon Europe **Findable, Accessible, Interoperable as well as Reusable (FAIR)**, to ensure its sound management. This means using standards and metadata to make data discoverable, specifying data sharing procedures and which data will be open, allowing data exchange via open repositories, as well as facilitating the reusability of the data. With that in mind, the following sections of the DMP lay out the methodology followed in the framework of MANOLO concerning making data findable, accessible, and interoperable as well as ensuring their preservation and open access, to increase its re-use.

3.1 Making data findable, including provisions for metadata

3.1.1 Data discoverability and identification mechanisms

MANOLO places special emphasis on enhancing the discoverability of the data collected/generated, or reused during its activities. **Open data produced during the implementation of the project are locatable employing a standard identification mechanism.** Indeed, MANOLO assigns globally resolvable **Persistent Identifiers (PIDs)** on any open data (more information on open data, as well as the respective repositories we plan on employing in the context of the project, are provided in section 3.2). An identifier is a unique identification code that is applied to a dataset so that it can be unambiguously referenced. For example, a catalogue number is an identifier for a particular specimen, and an ISBN code is an identifier for a particular book. PIDs are simply maintainable identifiers that allow for permanent reference to a digital object. In other words, PIDs are a way of giving digital resources, such as documents, images, and data records, a unique and persistent reference number.

At the same time, data that are not open are deposited in a searchable resource (i.e., the cloud web storage service of the project) and well-tailored identification mechanisms are utilized as well, in the form of standard naming conventions that will safeguard their consistency and make them easily locatable for partners within the frame of the project. Thus, the following subsection provides further analysis of naming conventions and versioning.

3.1.2 Naming conventions and versioning

Following a consistent set of naming conventions in the development of the project's data files can greatly enhance their searchability. With that in mind, MANOLO creates consistent data file names that provide clues to their content, status, and versioning, while also increasing their discoverability. In doing so, project partners as well as interested stakeholders can easily identify a file, as well as classify and sort them.

According to the UK Data Archive ([UK Data Service](#), 2017b), a best practice in naming conventions is to create brief yet meaningful names for data files that facilitate classification. The naming convention should avoid the utilisation of spaces, dots, and special characters (such as & or !), whereas the use of underscores is endorsed to separate elements in the data file name and make them understandable. At the same time, versioning should be a part of a naming convention to identify the changes and edits in a file.

With that in mind and to facilitate the reference of the datasets that are and will be produced during its implementation, MANOLO employs a standard naming convention that integrates versioning and considers the possibility of creating multiple datasets during an activity that entails data collection/generation. Indeed, MANOLO's naming convention considers this issue and addresses it by employing a unique element that captures the number of datasets that are produced under the same activity.

In particular, the **naming convention employed by the project is described below.**

[Name of project] _ [Name of Study] _ [Number of dataset] _ [Issue Date] _ [Version number]

- **Name of project:** MANOLO
- **Name of Study:** A short version of the name of the activity for which the dataset is created.
- **Number of dataset:** An indication of the number assigned to the dataset.
- **Issue Date:** The date on which the latest version of the dataset was modified (YYYYMMDD).
- **Version number:** The versioning number of a dataset.

With the above in mind, some **indicative examples** to showcase the naming structure applied in the context of MANOLO are provided below:

- **MANOLO_NeedsandChallengesOnlineSurvey_Dataset1_20240630_v1** – The first dataset was generated from the results of the online survey conducted to identify their perceived needs and barriers in the cloud-edge continuum landscape in terms of efficiency and trustworthiness, new technological solutions, key actors, and openly available resources. This is the first version of the dataset that was last modified on the 30th of June 2024 (30/06/2024).
- **MANOLO_EvaluationandValidation_Dataset2_20261231_v2** – The second dataset was created in the process of the evaluation and validation of the MANOLO solution under T6.3, presenting data and results from the validation procedure against the initial requirements and design principles. The last modification of this dataset, which in this case produced the second version of the dataset, was on the 31st of December 2025 (31/12/2026).

Versioning of information revises datasets uniquely identifiable and can be used to determine whether and how data changed over time, and to define specifically which version the creators/editors are working with. Moreover, effective data versioning enables understanding if a newer version of a dataset is available and which changes are between the different versions, allowing for comparisons and preventing confusion. In this context, **a clear version number indicator is used in the naming convention** of every data file produced during MANOLO to facilitate the identification of different versions.

3.1.3 Metadata allowing discovery

In addition to consistent naming conventions and versioning, the project also follows a metadata-driven approach to allow discovery and further increase the searchability of the data, while also facilitating its understanding and reuse. Metadata is defined as “data about data” or “information about information”¹⁰. It is usually structured textual information that describes the creation, content, or context of a digital resource – be it a single file, part of a single file, or a collection of many files. Metadata is the glue that links information and data across the World Wide Web. It is the tool that helps people to discover, manage, describe, preserve, and build relationships with and between digital resources¹¹.

In particular, three distinct types of metadata exist¹², as presented below:

- **Descriptive metadata**, used to identify and describe collections and related information resources. Descriptive metadata at the local level helps with searching and retrieving. In an online environment,

¹⁰ Huxley, L., & Jacobs, N. (2004). Online information services in the Social Sciences. Oxford: Chandos.

¹¹ Foulonneau, M., & Riley, J. (2008). Metadata for digital resources: Implementation, systems design, and interoperability. Oxford: Chandos.

¹² Caplan, P. (2003). Metadata fundamentals for all librarians. Chicago: American Library Association.

descriptive metadata helps to discover resources. Most of the time includes information such as the title, author, date, description, identifier, etc.

- **Administrative metadata** is used to facilitate the management of information resources. It is helpful for both short-term and long-term management and processing of data. This is information that will not usually be relevant to the public but will be essential for staff to manage collections internally. Such metadata may be location information, acquisition information, etc.
- **Structural metadata** enables navigation and presentation of electronic resources. It documents how the components of an item are organized. Examples of structural metadata could be how pages are ordered to form chapters of a book, a photograph that is included in a manuscript or a scrapbook, or the JPEG and TIF files that were created from the original photograph negative, linked together.

With that in mind, **data produced/used during MANOLO is discoverable with metadata** suitable to its content and format. The project employs **metadata standards** to produce rich and consistent metadata to support the long-term discovery, use, and integrity of its data. More details on the metadata standards adopted by MANOLO are provided in the following subsection.

3.1.4 Standards for Metadata Creation

MANOLO employs standards for creating metadata for data collected/generated by the project, to describe it with rich metadata and thus improve its discoverability and searchability. As a result, effective searching, improved digital curation, and easy sharing are realized. In addition, the metadata standards applied enable the integration of metadata from a variety of sources into other technical systems.

With that in mind, **for MANOLO' openly available data, the metadata standards provided by Zenodo are used**. Zenodo creates metadata to accompany the datasets that are uploaded to its repository, extending their reach to a wider audience of interested stakeholders. This metadata can be exported in several standard formats, including open and machine-readable ones (such as MARCXML, Dublin Core, and DataCite Metadata Schema), following the guidelines of OpenAIRE and is stored by Zenodo in JSON format according to a defined JSON schema¹³.

Project **data not available for reuse are also annotated with open and machine-readable metadata** following the **Dublin Core Metadata standard**. The Dublin Core Metadata element set (certified with the ISO Standard 15836) is a standard that can be easily understood and implemented, and as such, is one of the best-known metadata standards. It was originally developed as a core set of elements for describing the content of web pages and enabling their search and retrieval. Among the reasons for selecting this standard is also the fact that **Zenodo is compatible with Dublin Core metadata formats** and thus any initially closed data, that may become open at a later stage (e.g., due to a change in the consortium's policy), will not lose its metadata. With that said, the Dublin Core metadata standard is a simple yet effective set for creating rich metadata that will describe a wide range of resources. The fifteen-element "Dublin Core" described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the [Dublin Core Metadata Initiative \(DCMI\)](#). The full set of vocabularies also includes sets of resource classes, vocabulary encoding schemes, and syntax encoding schemes. **An online metadata generator is used** to produce the different metadata elements required (dublincoregenerator.com).

¹³ For more information on the JSON format and the JSON schema visit the following website: <http://json-schema.org/>

3.1.5 Search keywords included in the metadata

The project's data are provided with search keywords to optimize its findability as well as its ultimate reuse by interested stakeholders during its entire lifetime. With that in mind, the metadata standards employed by MANOLO provide opportunities for tagging the data collected/generated and its content with keywords. In general, keywords are a subset of metadata and include words and phrases used to name data. In the context of MANOLO, keywords are used to add valuable information to the data collected/generated as well as to facilitate the description and interpretation of its content and value.

Along these lines, the project's strategy on keywords is underpinned by the following principles:

- The who, the what, the when, the where, and the why should be covered.
- Consistency among the different keyword tags needs to be ensured.
- Relevant, understandable, and clear keywording ought to be sought.

In general, the keywords comprise terms related to trustworthy AI and the Cloud-Edge continuum. The keywords accurately reflect the content of the datasets and avoid words used only once or twice within them.

3.1.6 Offering metadata that can be harvested and indexed

We know that the wild diversity of the metadata accompanying open data across the plethora of online repositories (e.g., disciplinary archives, institutional repositories, open access journals) can serve as barriers to their findability and sharing amongst different research communities. This is why, in the context of MANOLO, we have aligned our metadata-creating approach with the **Open Archives Initiative (OAI)**, which promotes the use of a standard protocol for metadata harvesting, designed for better sharing and retrieval of data residing in distributed repositories. This protocol, namely the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH),¹⁴ promotes interoperability standards that facilitate the efficient dissemination of data amongst diverse communities¹⁵.

All structured **metadata linked to the project's open data are offered in a way that can be exported and harvested** via the OAI-PMH thanks to the standards we adopt for metadata creation (see section 3.1.4). The same standards also help us produce **metadata that facilitates indexing**. For instance, the use of the Dublin Core Metadata Standard (as further elaborated in section 3.3) provides a vocabulary of concepts with definitions in open-machine readable formats that enable easier indexing of metadata. Along these lines, several tools¹⁶ implement the Archives Initiative Protocol for Metadata Harvesting, such as **Arc source**, **EnhancedOAIserver**, and **eprints.org**, and can be used for harvesting our data by different repositories.

MANOLO's openly available data are uploaded to Zenodo, which is in line with FAIR principles, including the "To be Findable" principle. The metadata of each record uploaded to Zenodo is indexed and searchable directly in the Zenodo search engine immediately after publishing. The metadata of each record is sent to DataCite servers during DOI registration and indexed there.

3.2 Making data accessible

3.2.1 Repository

The **data produced by MANOLO and deemed open for sharing and reuse are deposited in and securely stored by Zenodo (www.zenodo.org)**, which constitutes an open data repository and has been specifically

¹⁴ Retrieved from: <https://www.openarchives.org/pmh/>

¹⁵ Corrado, E.M. (2005) 'The importance of open access, open source, and open standards for libraries', Issues in Science and Technology Librarianship.

selected to enable access to the project's open data free of charge. Zenodo builds and operates a simple service that enables researchers, scientists, EU projects, and institutions, among others, to share and showcase research results (including data and publications) that are not part of the existing institutional or subject-based repositories of the research communities. It accepts any file format, promotes peer-reviewed, openly accessible research, allows the creation of own collections, and is available free of charge both for MANOLO to upload and share data as well as for other stakeholders to explore, download and re-use this data.

Moreover, as a digital repository, Zenodo registers **Digital Object Identifiers (DOIs)** for all submitted data through DataCite¹⁶, which is the leading global non-profit organisation that provides PIDs (and specifically DOIs) for research data and preserves these submissions using the safe and trusted foundation of CERN's data centre, alongside the biggest scientific dataset in the world, the LHC's 100PB Big Data store¹⁷. This means that the data preserved in Zenodo will be accessible for years to come, and the DOIs will function as perpetual links to the resources. DOIs remain valuable since they are future-proofed against Uniform Resource Locator (URL) or even protocol changes, through resolvers (such as DOI¹⁸). With that in mind, an example of a DOI retrieved from this open repository follows the structure illustrated in Figure 1.

Figure 1: Typical DOI created by Zenodo

DOI 10.5281/zenodo.3901783

Accordingly, open-source **software** and related data will be deposited and securely stored on GitHub hosted by CeADAR (<https://github.com/manolo-euproject/>).

3.2.2 Data

Openly available and closed data

MANOLO follows the guidelines of the [Horizon Europe Data Management Plan Template](#), and is in line with FAIR principles and the rule “as open as possible, as closed as necessary”, aiming to “*make the data collected/generated openly available with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access*”¹⁹. This calls for project partners to disseminate the project's data that have the potential to offer long-term value to external stakeholders and do not harm the confidentiality and privacy of the stakeholders that contributed to the collection/generation of this data, maximising the beneficial impact of MANOLO.

Only anonymised and aggregated data are made open to ensure that data subjects cannot be identified in any report, publication, and/or dataset resulting from the project unless explicit consent has been collected following the respective GDPR provisions. The relevant project partner in each case **undertakes all the necessary anonymisation procedures** to anonymise the data in such a way that the data subject is no longer identifiable (more details on data management responsibilities are provided in Section 5.2).

To this end, it is important to keep in mind that during the process of data anonymisation, data identifiers need to be removed, generalised, aggregated, or distorted. Moreover, **anonymisation is different from pseudonymisation**, which falls under a distinct category in the GDPR - anonymisation theoretically destroys any way of identifying the data subject, while pseudonymisation allows for the data subject to be re-identified with additional information. Along these lines, the table which follows provides a **list of good practices** for the anonymisation of quantitative and qualitative data derived from the tour guide on data management of the Consortium of European Social Science Data Archives (CESSDA).

¹⁶ For more information on DataCite: <https://www.datacite.org/>

¹⁷ Retrieved from: <https://www.software.ac.uk/tags/zenodo>

¹⁸ Retrieved from: <http://dx.doi.org/>

¹⁹ Koulocheri, E. (2017). What is the Open Research Data Pilot? Retrieved from <https://www.openaire.eu/what-is-the-open-research-data-pilot>

Table 6: Good practices for data anonymisation

Type of data	Good practices
Quantitative data	- Remove or aggregate variables, or reduce the precision or detailed textual meaning of a variable. - Aggregate or reduce the precision of a variable such as age or place of residence. As a general rule, report the lowest level of geo-referencing that will not potentially breach respondent confidentiality. - Generalise the meaning of a detailed text variable by replacing potentially disclosive free-text responses with more general text. - Restrict the upper or lower ranges of a continuous variable to hide outliers if the values for certain individuals are unusual or atypical within the wider group researched.
Qualitative data	- Use pseudonyms or generic descriptors to edit identifying information, rather than blanking out that information. - Plan anonymisation at the time of transcription or initial write-up (longitudinal studies may be an exception if relationships between waves of interviews need special attention for harmonised editing). - Use pseudonyms or replacements that are consistent within the research team and throughout the project. For example, using the same pseudonyms in publications and follow-up research. - Use 'search and replace' techniques carefully so that unintended changes are not made, and misspelled words are not missed. - Identify replacements in the text clearly, for example with [brackets] or using XML tags such as <seg>word to be anonymised</seg>. - Create an anonymisation log (also known as a de-anonymisation key) of all replacements, aggregations, or removals made and store such a log securely and separately from the anonymised data files.

Source: Tour guide on data management of the CESSDA²⁰

With that in mind, the following table presents the data collected/generated during the project that are and will be made openly available. In case certain data cannot be shared (or need to be shared under restrictions), a justification for that choice is provided.

Table 7: Data Availability

No	Data	Availability	Notes
1	Desk research and interviews	Open	-
2	Survey's results	Open & Closed	Data collected from the online survey will be available only to the MANOLO consortium and the EU Commission. In case there is a need to share information for dissemination and communication purposes, GDPR requirements for personal data protection will be accounted for, and all personal information will be anonymised before being made openly available.

²⁰ Retrieved from: <https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/5.-Protect/Anonymisation>

No	Data	Availability	Notes
3	Data from the State of the Art and regulatory resources	Closed	Data collected will be available only to the MANOLO consortium and the EU Commission. In case there is a need to share information for dissemination and communication purposes, any sensitive information and identifiers related to the product development will be removed.
4	Memory, connection throughput, location of input and output data, energy efficiency	Closed	These data are closed, and it is used to implement effective management of the cloud edge continuum. It may be stored, but still in the domain of the MANOLO.
6	Use cases, datasets, and their derivatives	Closed	Private data from the use-case companies. Data generated from it is susceptible to being detrimental to these companies against their competitors, for being representative of their initial datasets.
5	Synthetic/Distilled dataset	Open	It will be based on public datasets.
7	Model results and statistics	Closed	-
8	LLM dataset and its generation results	Open	Dataset for LLM domain adaptation and responses from experiments.
9	T-Hyperparameter search on compression algorithms, with results and dataset + model metadata	Open	Benchmark datasets on Computer Vision and NLP are used to train models.
10	Curated texts for LLMs	Open	-
11	Generated responses from LLMs	Open	-
12	Resource topology: Nodes characteristics such as Architecture, CPU, Memory, Storage	Closed	These data are internally available only for the ML algorithm performing the continuum resource allocation.

No	Data	Availability	Notes
13	Use case AI applications description	Closed	These data are internally available only for the ML algorithm performing the continuum resource allocation.
14	Deployment information after optimization	Open	-
15	Electric power and energy consumption metrics consumed on bare metal hosts	Closed	-
16	Templates and Self-assessments	Open	Templates will be open and self-assessments will be used for publication. Templates show how to employ the Z-Inspection process in practice. Self-assessments validate the process and provide details on how MANOLO will establish Trustworthy AI.
17	Conversations	Closed	The goal of this data is to validate solutions generated in the project. This will only be available to the MANOLO consortium and the EU Commission.
18	Robot state information	Closed	
19	3D information of the environment	Closed	
20	2D information about the environment	Closed	
21	Model's performance in different nodes of the topology	Open	Performance baseline of the different models used in our use case, allowing for comparison before and after MANOLO.
22	Image classification, photograph tagging, and visual cues for predictions	Open	The goal of these data is to provide a benchmark for image classification across project tasks, which could become available for publication. Image classification datasets are assembled from existing publicly available datasets, with additional samples that can be added by partner contributions, provided they come with a proper license.
23	Face detection, Face retrieval.	Open & Close	The goal of these data is to provide a benchmark for image localization and feature extraction across project tasks. Face recognition datasets are both publicly available online and closed, the latter could be potentially used by the MANOLO consortium.
24	Sleep recordings for sleep stage decoding	Open	-

No	Data	Availability	Notes
25	Sleep recordings during real-time closed-loop auditory stimulation	Closed	-
26	Use cases, data evaluation, and validation	Open or closed, depending on the sensitivity of the data	Open or closed, depending on the sensitivity of the data. Evaluation and validation data will be available only to the MANOLO consortium and the EU Commission. If statistics are required to be shared, such data will be aggregated and anonymized before being made openly available, while personal data will be treated according to the GDPR.
27	Website and social media analytics	Open & Closed	Website analytics will be available only to the MANOLO consortium and the EU Commission. In cases where statistics were shared, data were aggregated and anonymized before being made openly available (e.g., reported in the publicly available Dissemination and Communication Plan of the project), while personal data were treated as expected by the GDPR.
28	Data collected from project events	Open & Closed	The personal data of participants will remain closed, as well as the confidential information of SMEs. Aggregated statistics may be published for promotion and reporting purposes.
29	Data collected from dissemination and communication actions	Open & Closed	Any personal information data will be aggregated/ anonymized before becoming openly available.
30	Newsletter subscriptions	Closed	Data from newsletter subscriptions will remain closed as it contains personal information and is useful only for internal reporting purposes.
31	Market analysis, Competition analysis, as well as financial planning	Closed	Data closely related to business planning may include sensitive financial information that cannot be made publicly available.
32	Constituting the EIAB	Open & closed	Data on the prospective and selected members were shared by Consortium members or were publicly available online. The following data are stored: • Name • Affiliation • Contact details (email) • Main expertise • Link to an online CV • Photography (for confirmed members only) Information about members who are not selected for the EIAB will be kept in order to keep a list of experts that could be contacted

No	Data	Availability	Notes
			at later stages of the project. This data will remain inside the Consortium. Prospective members who indicate that they are not interested in collaborating with the MANOLO project will be removed from the list.
33	Management of the EIAB	Open & closed	The status of the various documents (i.e., whether they will be made publicly available or confidential) will depend on the sensitivity of the document and/or the original data that the document refers to about the project.

It is important to note that **all personal data collected/generated are and will be considered as closed data prior to their anonymisation and aggregation** to safeguard the confidentiality of the data subjects.

More details about **the data related to the pilot studies** will be documented in **D8.4**, due in M36.

Data accessibility and availability

Public access to the open data is available through Zenodo, which automatically links to OpenAIRE, and GitHub. The data are fully accessible thanks to the included metadata and the search facility available on both online repositories. At the same time, closed data are and will be stored and shared amongst authorised members of the consortium through cloud storage and file-sharing providers, which constitute structures that maintain and manage data and make these data accessible over a network, usually the internet (i.e. ARX.NET cloud repository). Before starting to use these cloud services from providers situated both inside and outside the European Economic Area (EEA), we have ensured that they comply with the relevant GDPR requirements.

The following table presents where data are and will be made accessible in the context of MANOLO.

Table 8: Data Accessibility

No	Data	Accessibility
1	Desk research and interviews	Zenodo
2	Survey's results	Zenodo
3	Data from the State of the Art and regulatory resources	Closed data
4	Memory, connection throughput, location of input and output data, energy efficiency	Closed data
5	Use cases datasets and their derivatives	Closed data
6	Synthetic/Distilled dataset	Zenodo and GitHub
7	Model results and statistics	Closed data
8	LLM dataset and its generation results	They will be referenced as they are publicly available. TBD

No	Data	Accessibility
9	T-Hyperparameter search on compression algorithms, with results and dataset + model metadata	They will be referenced as they are publicly available. TBD
10	Curated texts for LLMs	They will be referenced as they are publicly available. TBD
11	Generated responses from LLMs	They will be referenced as they are publicly available. TBD
12	Resource topology: Nodes characteristics such as Architecture, CPU, Memory, Storage	Closed data
13	Use case AI applications description	Closed data
14	Deployment information after optimization	Zenodo & GitHub
15	Electric power and energy consumption metrics consumed on bare metal hosts	Closed data
16	Templates and self-assessments	Templates: MANOLO Website Self-assessments: Closed data
17	Conversations	Closed data
18	Robot state information	Closed data
19	3D information of the environment	Closed data
20	2D information of the environment	Closed data
21	Model's performance in different nodes of the topology	Zenodo
22	Image classification, photograph tagging, and visual cues for predictions	Zenodo
23	Face detection, face retrieval	Zenodo for open data, password-protected ARX.NET data repository for closed data.
24	Sleep recordings for sleep stage decoding	The dataset has been published in openNeuro and will also be uploaded to Zenodo by the end of 2025.
25	Sleep recordings during real-time closed-loop auditory stimulation	Closed data for now. We cannot provide this information yet. We will evaluate whether we can share the data and how.
26	Use cases, data evaluation, and validation	Zenodo (related to open data) and proprietary software (e.g., LAU's OneDrive) for sensitive data
27	Website and social media analytics	Closed data
28	Data collected from project events	MANOLO Website, Zenodo

No	Data	Accessibility
29	Data collected from dissemination and communication actions	MANOLO Website, Zenodo
30	Newsletter subscriptions	Closed data
31	Market analysis, Competition analysis, as well as financial planning	Closed data
32	Constituting the EIAB	MANOLO website, MANOLO social media
33	Management of the EIAB	MANOLO website

Restrictions on use

By utilising Zenodo for sharing the project's openly available data, MANOLO can apply **different levels of accessibility** for this data, considering any relevant issues (such as ethics, rules of personal data, intellectual property, commercial, privacy-related, security-related, etc.).

More specifically, **Zenodo offers the following levels of data accessibility**:

- **Open access:** Data remains available for re-use. Nevertheless, the level at which this data can be re-used is also determined by their accompanied licence for re-use (see subsection 3.4.3).
- **Embargoed status:** Access to the data will be restricted until the end of the embargo period, at which time, the content will automatically become publicly available.
- **Restricted access:** The data will not be made publicly available, and sharing will be made possible only by the approval of the project partner that has the responsibility for the data.
- **Closed access:** The data are protected against unauthorized access at all levels, and only members of the consortium have the right to access it.

The same principle applies also for GitHub hosted by CeADAR, within which **private and public repositories** can be created with differing sharing and access options.

Accordingly, the data repository that is used for the needs of MANOLO, hosted by ARX.NET, allows both publicly available access (for publicly shared data) or private use only to project partners through password-protected access.

Project partners mainly use the open access level to disseminate the project's data among the interested stakeholders. Data that are not available for reuse are accessible only by authorised partners of the MANOLO consortium and /or authorised personnel from the funding authority of the project.

Moreover, **MANOLO ensures open access to all peer-reviewed scientific publications** that may be produced in the framework of the project. In particular, according to the Grant Agreement, MANOLO will:

- At the latest at the time of publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a trusted repository for scientific publications.
- Ensure immediate open access to the deposited publication — via the repository — under the latest available version of the Creative Commons Attribution International Public License (CC BY) or a license with equivalent rights. Moreover, for monographs and other long-text formats, the license may exclude commercial uses and derivative works (e.g., CC BY-NC, CC-BY-ND).
- Ensure information is given, via the repository, about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

Along these lines, this section has provided the methodology applied in the frame of MANOLO to ensure that its data are as openly accessible as possible by any stakeholder who may find it interesting for reuse. In this context, MANOLO also focuses on providing metadata standards and appropriate metadata vocabularies to increase data interoperability. The following section provides further details in this respect.

It should be noted here that INRIA scientific publications are automatically available on the open Library HAL²¹. In addition, PAL upload scientific publications to arXiv²².

Identity ascertainment and data access committee

The identity of stakeholders who want to access the open data on Zenodo, GitHub, or ARX.NET's database does not necessarily have to be ascertained, as the uploaded data are publicly open and no authorization is needed. On the other hand, closed-for-the-public data are available only to authorized consortium partners through dedicated mechanisms provided by the cloud storage service employed by the respective partners to deposit the data. As further elaborated in Section 6 of this DMP, technical access controls are built into the MANOLO repositories, to ascertain the identity and access rights of those who want to access the data.

The need for a data access committee to evaluate or approve access requests to personal data is not foreseen because only authorized partners have access to the project's closed data, accessible only by using their credentials (username/password), and no third-party will re-use them for their benefit.

3.2.3 Metadata

Availability and licences

Metadata of deposited publications generated in the context of MANOLO are **open under a Creative Commons Public Domain Dedication (CC 0)** or equivalent, in line with the FAIR principles for data management adopted by the project (in particular machine-actionable). Such **metadata provides information, at least, about the following**:

- The publication at hand (author(s), title, date of publication, publication venue);
- Reference to the Horizon Europe funding;
- The name of the project, including its acronym and Grant Agreement number;
- Any particular licensing terms which may apply (depending on the chosen license);
- Persistent Identifiers (PIDs) that have been attributed to the publication;
- Authors involved in the action, their organisations, and the project itself.

Where applicable, the metadata also includes PIDs for any research output, or any other tools and instruments needed to validate the conclusions of the publication. The metadata is available through Zenodo and GitHub. It is quite unlikely that Zenodo or GitHub will terminate its operation and stop providing its services, but in such a case, all data, metadata, code, and documentation uploaded will be transferred and hosted to other suitable repositories without undue delay. In this respect, it is important to note that, since all of MANOLO' openly available data make use of PIDs (i.e., DOIs), which is further elaborated in subsection 3.1.1, the links to the data will not be affected. In parallel, the project's data that are not openly available for sharing are deposited, together with their accompanying metadata, code, and documentation (if necessary), to the cloud web storage service employed by the project.

²¹ <https://hal.archives-ouvertes.fr>

²² <https://arxiv.org/>

Methods, software tools, and documentation to access the data

MANOLO emphasises the accessibility of the data collected/generated during the project. With that in mind, no specialised method, software tool, and/or documentation is expected to be needed at the moment to access the data. Stakeholders have the ability to access the data by simply using their web browser (e.g., Mozilla, Google Chrome, Edge, Safari, etc.) through their computers (either desktop or laptop), smartphones, and/or tablets.

More specifically, through the MANOLO website, clear links to the selected repositories allow direct navigation to the publicly available datasets in all three online repositories (i.e., Zenodo, GitHub, ARX.NET's repository). For Zenodo and GitHub, stakeholders are also able to utilise the search engine of the repositories to search for interesting data. By typing the name of the project (or any other relevant keyword connected to the MANOLO data), the search engine will direct the user to the project's data, ready to be explored and reused. Moreover, since the data are available in open formats, we are ensuring that they can be appropriately read by a range of different software that is widely and freely accessible to all potential users of the data.

Closed data are only accessed by authorised project partners through the use of a cloud storage service. Again, no specialised method, software tool, and/or documentation is needed for this end.

3.3 Making Data Interoperable

Data interoperability refers to the ability of systems and services that create, exchange, and use data to have clear, shared expectations for the contents, context, and meaning of that data²³. With that in mind, MANOLO has adopted in its data management methodology the use of metadata vocabularies, standards, and methods that increase the interoperability of the data collected/generated through its activities.

More specifically, **the interoperability of the data that are not publicly shared is facilitated by the use of the Dublin Core Metadata standard**. This standard is a small “metadata element set” that accounts for issues that must be resolved to ensure that data meets traditional standards for quality and consistency while remaining broadly interoperable with other data sources in the linked data environment. The fifteen elements of the standard provide a vocabulary of concepts with natural-language definitions (e.g., title, creator, author, etc.) that are instantly converted into open machine-readable formats (such as XML, HTML, etc.), enabling machine-processability. Each element is optional and may be repeated, while the standard itself offers ways to refine them, encouraging the use of encoding and vocabulary schemes. The vocabulary of the Dublin Core Metadata standard is presented in the following table²⁴.

Table 9: Dublin Core Metadata standard vocabulary

No	Element	Element definition
1	Title	A name given to the resource.
2	Creator	The entity is primarily responsible for making the content of the resource.
3	Subject	The topic of the content of the resource.
4	Description	An account of the content of the resource.
5	Publisher	An entity responsible for making the resource available.

²³ L. Steele & T. Orrell (2017). The frontiers of data interoperability for sustainable development. Publish What You Fund and Development Initiatives

²⁴ Sugimoto, S., Baker, T., & Weibel, S. L. (2002). Dublin Core: Process and Principles. Lecture Notes in Computer Science Digital Libraries: People, Knowledge, and Technology, 25-35.

No	Element	Element definition
6	Contributor	An entity responsible for making contributions to the content of the resource.
7	Date	A date associated with an event in the life cycle of the resource
8	Type	The nature or genre of the content of the resource.
9	Format	The physical or digital manifestation of the resource.
10	Identifier	An unambiguous reference to the resource within a given context.
11	Source	A reference to a resource from which the present resource is derived.
12	Language	A language of the intellectual content of the resource.
13	Relation	A reference to a related resource.
14	Coverage	The extent or scope of the content of the resource.
15	Rights	Information about rights held in and over the resource.

Along similar lines, **the interoperability of openly available data is facilitated through Zenodo** since its metadata is stored internally in JSON format according to a defined JSON schema. This encloses HTML microdata that allows machine-readable data to be embedded in HTML documents in the form of nested groups of name-value pairs. Moreover, the JSON schema provides a collection of shared vocabularies in microdata format that can be used to markup pages in ways that can be understood by the major search engines. Moreover, all metadata linked to the open data are exported via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and can be harvested. The OAI-PMH develops and promotes interoperability standards that facilitate the efficient dissemination of data amongst diverse communities²⁵.

Similarly, GitHub can facilitate storing metadata either through Git notes²⁶ or by associating a description with a branch. The metadata stored can follow the same JSON format.

3.4 Increase data reuse

3.4.1 Documentation for validating data analysis and facilitating data re-use

By utilising Zenodo for sharing the project's openly available data, MANOLO ensures the facilitation of data access, validation, and reuse, in compliance with the general policies of Zenodo regarding content, access, and reuse. More specifically, the following principles are followed by Zenodo to make data re-useable according to the FAIR principles:

- **R1: (meta)data are richly described with a plurality of accurate and relevant attributes**

Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.

- **R1.1: (meta)data are released with a clear and accessible data usage license**

²⁵ Corrado, E.M. (2005) 'The importance of open access, open source, and open standards for libraries', Issues in Science and Technology Librarianship.

²⁶ <https://git-scm.com/docs/git-notes>

License is one of the mandatory terms in Zenodo's metadata and refers to an Open Definition license. Data downloaded by the users are subject to the license specified in the metadata by the uploader.

- **R1.2: (meta)data are associated with detailed provenance**

All data and metadata uploaded are traceable to a registered Zenodo user. Metadata can optionally describe the original authors of the published work.

- **R1.3: (meta)data meet domain-relevant community standards**

Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

On the other hand, by utilising GitHub for sharing the project's openly available data, and specifically the open-source solutions developed along the datasets used for the various experiments, MANOLO offers access, and thus increased reuse, to the scientific community that uses it as their main knowledge sharing online repository.

3.4.2 License schemes to permit the widest use possible

The application of a licence to MANOLO open data is a simple way to ensure that any interested third party can reuse it. In this context, licences are the instrument that permit a third party to copy, distribute, display, and/or modify the project's data only for the purposes that are set by the licence. Licences typically grant permissions on the condition that certain terms are met. While the precise details vary, three conditions are commonly found in licences, which are attribution, non-derivative, and non-commerciality.

Along these lines, MANOLO publishes openly available data under the **Creative Commons licensing scheme** to foster their reuse and build an equitable and accessible environment for them. Zenodo provides MANOLO the **opportunity to publish its open data under five Creative Commons licences** as follows:

- **Creative Commons Attribution-Share Alike 4.0 (CC BY-SA 4.0)**, according to which any third party can freely copy, distribute, display, and modify the datasets for any purpose. Remix, transform, or build upon data, must be distributed under the same license as the original. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.
- **Creative Commons Attribution 4.0 International (CC BY 4.0)**, according to which any third party can freely copy, distribute, display, and modify the datasets for any purpose. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.
- **Creative Commons Attribution-No Derivatives 4.0 International (CC BY-ND 4.0)** under which any third party can freely copy, distribute, display, and modify the datasets for any purpose. Remix, transform, or build upon data, however, it must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.
- **Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)**, based on which third parties can copy, distribute, display, and modify the datasets for any purpose other than commercial, unless they get permission from project partners first. Third parties must give

Figure 4: CC BY-SA 4.0



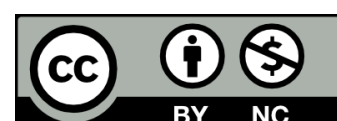
Figure 7: CC BY 4.0



Figure 10: CC BY-ND 4.0



Figure 13: CC BY-NC 4.0



appropriate credit, provide a link to the license, and indicate if changes were made.

- **Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International** (CC BY-NC-ND 4.0), according to which third parties can copy, distribute, display, and modify the datasets for any purpose other than commercial unless they get permission from project partners first. Remix, transform, or build upon data, however, it must not be distributed. Third parties must give appropriate credit, provide a link to the license, and indicate if changes were made.

Figure 16: CC BY-NC-ND 4.0



Beyond Zenodo licence schemes, GitHub offers a wider superset of licenses²⁷ that extends to other types of licenses besides Creative Commons, including but not limited to Academic Free License v3.0, Berkeley Software Distribution 2 and 3 licences, European Union Public License 1.1, GNU General Public License v2.0/v3.0, Open Software License 3.0, Educational Community License v2.0, and more.

Different licensing schemes may be selected to better fit the needs of MANOLO's open data, ensuring not only their long-term preservation and reuse but also the interests of the consortium along with the rights of individuals for whom the data is about. In such a case, this subsection of the DMP will be updated accordingly.

3.4.3 Availability for re-use

The reuse of data is a key component of MANOLO's methodology for making data FAIR. Making data available for reuse ensures interested stakeholders, other than project partners, can benefit from this data, contributing towards maximising the impact of the project. Rich metadata is created based on metadata standards that enable proper discovery as well as appropriate licensing schemes, facilitating the reuse of MANOLO's open data, allowing them to find valuable utility.

In principle, it is expected that data will become available for reuse no later than 120 days after the end of its processing in the framework of the project (i.e., collection, anonymisation, aggregation, etc.) to ensure that any additional data management activities required to this end do not compete with the timely delivery of the project's planned outputs.

With that in mind, the expected time that MANOLO's data are made openly accessible and uploaded to Zenodo is indicatively provided in the following table:

Table 10: Expected time that the data will be made open

No	Data	Expected time for making the data open	Notes
1	Desk research and interviews	Published	10.5281/zenodo.13143231 10.5281/zenodo.13143347
2	Survey's results	Published	10.5281/zenodo.13143375 10.5281/zenodo.13143387
3	Data from the State of the Art and regulatory resources	- (closed data)	No public sharing will be available
4	Memory, connection throughput, location of input and output	- (closed data)	No public sharing will be available

²⁷ <https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/licensing-a-repository>

No	Data	Expected time for making the data open	Notes
	data, energy efficiency		
5	Use cases datasets and their derivatives	- (closed data)	-
6	Synthetic/Distilled dataset	By the end of the project or when related milestones (publications) are reached.	-
7	Model results and statistics	- (closed data)	No public sharing will be available
8	LLM dataset and its generation results	Published	https://huggingface.co/datasets/yahma/alpaca-cleaned
9	T-Hyperparameter search on compression algorithms, with results and dataset + model metadata	Published	https://huggingface.co/datasets/yahma/alpaca-cleaned
10	Curated texts for LLMs	Published	https://huggingface.co/datasets/yahma/alpaca-cleaned
11	Generated responses from LLMs	Published	https://huggingface.co/datasets/yahma/alpaca-cleaned
12	Resource topology: Nodes characteristics such as Architecture, CPU, Memory, Storage	- (closed data)	No public sharing will be available
13	Use case AI applications description	- (closed data)	No public sharing will be available
14	Deployment information after optimization	31/12/2026	Upon T4.4 completion
15	Electric power and energy consumption metrics consumed on bare metal hosts	- (closed data)	No public sharing will be available
16	Templates and self-assessments	Templates M35 at the latest. Self-assessments (closed)	Upon T.5.5 completion

No	Data	Expected time for making the data open	Notes
17	Conversations	- (closed data)	No public sharing will be available
18	Robot state information		
19	3D information of the environment		
20	2D information of the environment		
21	Model's performance in different nodes of the topology	By 31/12/2026	-
22	Image classification, photograph tagging, and visual cues for predictions	"v1.0: CW24, 12 Jun 2025, Final Version: CW40, Oct 2025"	-
23	Face detection, face retrieval	N/A (for closed data)	Additional face recognition data that might be used for the needs of the project will not be made public.
24	Sleep recordings for sleep stage decoding	Published	https://openneuro.org/datasets/ds005555/versions/1.1.0 This dataset will be uploaded to Zenodo by the end of 2025.
25	Sleep recordings during real-time closed-loop auditory stimulation	2026 at the earliest	Closed data for now. This might be re-evaluated.
26	Use cases, data evaluation, and validation	31/12/2026	Upon T6.3 completion
27	Website and social media analytics	- (closed data)	No public sharing will be available
28	Data collected from project events	31/12/2026	Upon T7.1 completion
29	Data collected from dissemination and communication actions	31/12/2026	Upon T7.1 completion
30	Newsletter subscriptions	- (closed data)	No public sharing will be available
31	Market analysis, Competition analysis, as well	- (closed data)	No public sharing will be available

No	Data	Expected time for making the data open	Notes
	as financial planning		
32	Constituting the EIAB	Published	https://manolo-project.eu/advisory-board/ Continuous updates throughout the project
33	Management of the EIAB	31/12/2026 at the latest	-

3.4.4 Data provenance

Data provenance is the documentation of where a piece of data comes from and the processes and methodology by which it was produced. Put simply, provenance answers the questions of why and how the data was produced, as well as where, when, and by whom²⁸. Accurately recording data provenance is a cornerstone of good data management. MANOLO uses specific elements of the **Dublin Core Metadata Standards**²⁹ and the **W3C Provenance Data Model**³⁰, to generate specific text files (e.g., README) that accurately capture the history of each data entity throughout its versions (e.g., based on the DOI versioning Zenodo provides)³¹.

3.4.5 Data Quality Assurance Processes

Quality Assurance (QA) and **Quality Control** (QC) activities³² are an integral part of MANOLO' data management methodology and are implemented before the publication of any data to online repositories (e.g., Zenodo, GitHub, etc.), safeguarding the transparency, consistency, comparability, completeness, and accuracy of the data.

QA is a planned system of review procedures conducted outside the framework of developing a dataset by personnel not directly involved in the dataset development process. In the context of MANOLO, it takes the form of **peer reviews of methods and/or data summaries** to assess the quality of the dataset and identify any need for improvement, ensuring that the dataset correctly incorporates the scientific knowledge and data generated.

QC is defined as a system of checks to assess and maintain the quality of the dataset being compiled. The relevant procedures of MANOLO are designed to provide routine technical checks as they measure and control data consistency, integrity, correctness, and completeness, as well as identify and address errors and omissions. In this context, QC checks cover everything from data acquisition and handling to the application of approved procedures and methods, and documentation. Some of the general quality checks undertaken in the framework of the project include checking (i) for transcription errors in data input; (ii) that scale measures are within the range of acceptable values; and (iii) whether proper naming conventions are used.

²⁸ <https://ardc.edu.au/resource/data-provenance/>

²⁹ https://www.dublincore.org/resources/userguide/creating_metadata/#Provenance

³⁰ <https://www.w3.org/TR/prov-dm/>

³¹ <https://help.zenodo.org/>

³² 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Vol. 1 General Guidance and Reporting, CHAPTER 6 Quality Assurance / Quality Control and Verification.

4. Other research outputs

The implementation of MANOLO entails activities that generate or reuse other research outputs. This includes digital (e.g., software, workflows, protocols, models, etc.) research outputs, besides the data described in the second chapter of this DMP. The management of these other research outputs in the context of MANOLO is in line with FAIR data principles, as concisely outlined in the table that follows.

Table 11: Other digital or physical research outputs

Work Package / Task	Research output	Brief Description	Type	Expected size	Interested stakeholders and benefits	Availability	Accessibility (repository)	Expected time for making them open
WP2 / WP3	Academic contributions and dissemination.	Publications, conference transcripts/slides, and dissemination material.	.pdf, .docx, .pptx	Several MB	CeADAR, MANOLO (for metrics), and researchers or potential customers.	Open	Materials or relevant links to materials can be compiled on the MANOLO website	On the fly (after they are approved by the academic community).
WP3	AI models	AI models were trained with the different datasets available for the project.	Pickled files	Several GB in total	MANOLO partners, as the models have different dimensions, from AI development to trustworthiness and ethics.	Closed	-	-
WP3/T3.3	Paper and Web Service	The findings of T3.3 will be presented in a research paper, as well as the D3.1 v1 and v2. Furthermore, access to the algorithms and methods implemented as	.pdf and REST API (JSON)	-	1. Technology providers, commercial/industry stakeholders, 2. Academia & Researchers, 3. Governmental/policy stakeholders: These three stakeholder groups	The code for the “HW-aware Neural Architecture Search” of T3.3 will not be open-sourced, but the findings based on this	According to the standards of D3.1	Deadline for D3.1 v1: M18 → June 2025 Deadline for D3.1 v2: M30 → June 2026

Work Package / Task	Research output	Brief Description	Type	Expected size	Interested stakeholders and benefits	Availability	Accessibility (repository)	Expected time for making them open
		part of 3.3 will be provided via web services. The API provides a report of the Pareto front of several optimized NNS KPIs.			will benefit from the novel knowledge presented in the deliverables based on the activities of T3.3. These stakeholders can utilize these findings to focus their technological roadmaps and research.	framework will be published according to the standards of D3.1 and papers.		
WP3/ Task 3.4	We will develop an open-source library.	This library is meant to grow neural network architectures in a way that adapts to the machine learning task at hand.	Repository on GitLab.	-	Partners interested in developing frugal neural architectures might want to test it.	The library will be open source. Partners are free to use this library for academic purposes. For industrial or commercial uses, licences may apply. Check with us in such cases.	https://gitlab.inria.fr/mverbock/tinypub	The library pre-existed the MANOLO project but will be further developed thanks to Task 3.4.
WP4 and Task 4.4	Model for optimized allocation	An integrated system of model and learning will be allocated in the continuum for the AI/ML models to be	Digital (Integrated files including API,	7435 KB	After completing the simulation of allocating AI/ML models to the cloud-edge continuum, system	Closed	GitLab	30/09/2026, Upon T4.4 completion

Work Package / Task	Research output	Brief Description	Type	Expected size	Interested stakeholders and benefits	Availability	Accessibility (repository)	Expected time for making them open
		deployed in the continuum optimally	JSON, .py)		integration stress testing in a lab environment will be conducted to evaluate the performance of the models in the hardware. After that, it will be made available for functional integration into the use cases for the other partners.			
WP5 and Task 5.1	Software	Extend the KOBE Open Benchmarking Engine to apply ML workloads to systems with heterogeneous hardware components and dynamic network connectivity	Text	Small	-	-	-	-
WP6/T6.3	Report	A report on lessons learned and best practices for designing, developing, and deploying trustworthy AI will be drafted and	Digital and physical (.docx or .pdf formats)	-	From the point of view of the European Commission and the general public, findings from evaluation and validation activity	Open	MANOLO website	31/03/2027, Upon T6.3 completion and approval

Work Package / Task	Research output	Brief Description	Type	Expected size	Interested stakeholders and benefits	Availability	Accessibility (repository)	Expected time for making them open
		disseminated based on evaluation and validation outcomes.			will be used to form conclusions such as best practices for designing and deploying trustworthy efficient AI solutions in the cloud-edge continuum, as well as recommendations on how to best account for ethical aspects in policy frameworks to foster adoption in a way that respects human rights.			

5. Allocation of resources

5.1 Estimated costs for making data FAIR

The costs required for making the data collected/generated or reused during MANOLO's activities FAIR are integrated into the budget of the project. With that in mind, the table that follows provides an overview of the estimated costs of making data FAIR, as well as their budget source within the framework of MANOLO.

Table 12: Estimated costs for making data FAIR

No	Data Processing / Management Activity	Budget source	Total estimated effort in Person Months ³³	Total estimated cost in Euros ³⁴
1	Collection	Budget allocated to the WP under which the respective data are processed	36,09	217.550,80 €
2	Documentation	Budget allocated to the WP under which the respective data are processed	10,31	62.157,37 €
3	Storage	Budget allocated to the WP under which the respective data are processed	5,16	31.078,69 €
4	Access and security	Budget allocated to the WP under which the respective data are processed	5,16	31.078,69 €
5	Preservation	Budget allocated to the WP under which the respective data are processed	2,58	15.539,34 €
6	Availability and re-use	Budget allocated to the WP under which the respective data are processed	15,47	93.236,06 €
7	Overall data management	WP8	3,65	21.995,63 €
Total			78,40	472.636,56 €

To produce the estimations of the costs for making data FAIR in the context of MANOLO, a series of **assumptions** were made, taking into account the respective **guidelines** provided by the Research Data Management Support, a multidisciplinary network of data experts within Utrecht University³⁵, as well as of the UK Data Service and its data management costing tool³⁶. With that in mind, the estimated costs for making MANOLO's data FAIR cover **data-related activities and resources across the data lifecycle**, spanning from collection and documentation through storage and preservation over to sharing and re-use.

³³ The total estimated effort for each data processing/management activity reflects the cumulative effort for the implementation of the respective activity for all data collected/generated across the different WPs of MANOLO.

³⁴ The total cost of each data processing/management activity is calculated by multiplying the effort estimated for the respective activity with the weighted average cost of a person per month in the framework of MANOLO.

³⁵ Research Data Management Support. Guides: Costs of data management. Utrecht University. Retrieved from: <https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management>

³⁶ UK Data Service. Costing Data Management. Retrieved from: <https://www.ukdataservice.ac.uk/manage-data/plan/costing>

In particular, costs for **data collection** cover activities necessary for acquiring external datasets (if required), gathering/generating new data, transcribing (if applicable), formatting and organising this data, as well as acquiring informed consent from data subjects. This data processing activity reflects the majority of the costs required for making data FAIR, as the majority of MANOLO's data constitutes new data collected/generated throughout the project. At the same time, **data documentation** costs address the effort required for describing data (e.g., marking data with variable and value labels, code descriptions, etc.) as well as creating well-defined metadata along with a meaningful description of the context and methodology of how data were collected/generated and processed (where necessary).

Costs for **data storage** include the resources required for ensuring adequate storage space for the data as well as the effort necessary for conducting data back-ups, while **data access and security** costs encompass costs related to ensuring access to the data as well as protecting it from unauthorised access or use from disclosure. Given that the storage of MANOLO's data do not require the procurement of additional space (other than what is already available to project partners) as well as that no special measures or software are required to access and secure the data (other than what is inherently built into the repositories of MANOLO's data), such costs are kept to a minimum.

Data preservation costs, on the other hand, are estimated relatively higher than data storage, access, and security costs, as additional effort is required in several cases to convert the collected/generated data from their original form (e.g. physical interview transcripts) to an open and/or machine-readable format suitable for long-term preservation (e.g. to a .xlsx format.). Adequate effort for **data availability and reuse** costs is also foreseen to safeguard the appropriate digitisation and anonymisation of the data, as well as cover any resources required for data sharing and cleaning. Along the same lines, the appropriate effort is foreseen for **overall data management** as well, to cover the effort related to the operationalisation of data management in the framework of MANOLO.

Finally, costs for **long-term preservation** in the framework of MANOLO are assumed to be negligible since the open data of the project will be hosted in the repository of Zenodo free of charge.

5.2 Data management responsibilities

For the effective, proper, and secure handling of the data collected/generated or reused in the frame of MANOLO, specific data management roles have been established within the data management methodology and procedures of the project. These responsibilities are outlined in this section of the DMP and are as follows.

Project Coordinator (PC) and Project Management Office (PMO): responsible for overall data management in the framework of MANOLO, including the elaboration of the DMP and its updates (when necessary, along with the support of all partners). At the same time, the PC and PMO are responsible for the elaboration of proper templates for the Informed Consent Form and the Data Subject Request Form to be appropriately adjusted and utilised by project partners during the relevant activities of the project as well as for drafting the project's Privacy Policy that has been uploaded on the project's website. The PC and PMO, in collaboration with the relevant project partners (e.g., Task Leaders), examine if additional specific privacy policies are required for certain project tasks and coordinate the elaboration of such privacy policies. Finally, the PC and PMO coordinate with Work Package Leaders and Co-leaders, Task Leaders, and Responsible Partners to determine whether and how the data collected/generated or reused by the project is shared and becomes available for reuse, contributes to its quality assurance, and uploads the project's openly available data to Zenodo.

Work Package Leaders (WPL): The WPL is responsible for coordinating the implementation of the data processing activities performed under the WPs they are leading. Moreover, they align with the PC and the respective Work Task Leader on whether and how the data gathered/produced under the tasks that fall within

the WP they are leading are shared and/or reused. This includes the definition of access procedures as well as potential embargo periods, along with any necessary software and/or other tools that may be required for data sharing and re-use. Finally, the WPLs are mainly responsible for assuring the quality of the data stemming from the activities of the WP they are leading, including assessing their quality and indicating any need for improvement to the respective Work Task Leaders.

Work Task Leaders (WTL): WTLs are responsible for the data collected/generated in the frame of the tasks that fall under their leadership, as well as for safeguarding their appropriate and timely processing. Moreover, they are responsible for properly adjusting the Informed Consent Form and Data Subject Request Form templates to the needs and specificities of the activities carried out in the task they are leading. WTLs are responsible for identifying the need for a specific privacy policy regarding the task they are leading and collaborating with the PC for drafting and releasing it to the public. Finally, they undertake any necessary actions to prepare the data collected/generated or reused through the tasks they are leading for sharing either within the consortium or openly (including the use of proper naming conventions, application of suitable anonymisation techniques, creation of appropriate metadata and documentation, etc.).

Partners: All project partners (including Affiliated Entities) are tasked to collect, digitise, anonymise, store, destroy and/or otherwise process data for the specific purpose of the activity in which they have been collected/generated within the project. They are responsible for appropriately collecting the necessary consent for processing data as well as for ensuring that the Informed Consent Form and the Data Subject Request Form used to this end are properly adjusted to the needs of the activity they are participating (including references to the project's Privacy Policy and any other applicable specific privacy policies) and, in any particularities, applicable to their organisation while ensuring adherence to provisions of relevant national data protection legislation in their respective country. Moreover, they are responsible for managing the consent they have collected, intending to demonstrate their compliance with the relevant applicable EU and national regulation(s). Finally, they perform quality checks to assess and maintain the quality of the dataset(s) held within their records.

Data repositories: Data repositories are tasked with the storage and long-term preservation of the project's data. In this respect, Zenodo and GitHub maintain and preserve the openly available data of MANOLO, enabling its sharing and re-use. To this end, Zenodo assigns metadata and DOIs to the data, while also taking all necessary measures to securely back up the data and restore it, safeguarding its long-term preservation. Accordingly, metadata is included in GitHub, where state-of-the-art security is ensured³⁷. Large datasets are stored at ARX.NET 20GB disk array (the space of the disk array might be reassessed depending on the final size of the datasets stored) and become either publicly available or available to project partners through password-protected access. The public datasets are linked from the MANOLO website through a data rate throttling mechanism.

In this context, the following table illustrates the allocation of data management responsibilities amongst the members of the MANOLO consortium per data collected/generated under each WP.

³⁷ <https://github.com/security>

Table 13: Data management responsibilities of MANOLO' partners per data collected/generated under each WP

WP	WPL	Data	Tasks	WTL	Responsible Partners
1	NUIDUCD-CeADAR	Desk research and interviews	T1.1	Q-PLAN	All partners
		Survey's results			
		Data from the State of the Art and regulatory resources			
		Memory, connection throughput, location of input and output data, energy efficiency	T1.3	ATOS IT	NUIDUCD-CeADAR, UPC, TUBS, NCSR "D", FDI, Fraunhofer, INRIA, ARCADa, PAL ROBOTICS, Bit&Brain, ARX.NET S.A., Q-PLAN, EIT DIGITAL and UPSaclay
2	NCSR "D"	Use cases datasets and their derivatives	T2.2	NCSR "D"	NUIDUCD-CeADAR, UPC, ATOS IT, EVIDEN RO, FDI, INRIA, ARCADa, PAL ROBOTICS, Bit&Brain, ARX.NET S.A.
		Synthetic/Distilled dataset	T2.3	NUIDUCD-CeADAR	NCSR "D", FDI, Fraunhofer, INRIA
		Use cases datasets and their derivatives			
3	Fraunhofer	Model results and statistics	T3.1	NUIDUCD-CeADAR	FDI, Fraunhofer, INRIA
		T-Hyperparameter search on compression algorithms, with results and dataset + model metadata	T3.2		FDI, Fraunhofer, INRIA, UPSaclay, ARCADa, KU Leuven, LAU, EIT DIGITAL
		LLM dataset and its generation results			
4	UPC	Resource topology: Nodes characteristics such as Architecture, CPU, Memory, Storage	T4.2	TUBS	UPC, ATOS IT
		Use case AI applications and their requirements			
		Deployment information after optimization	T4.4	UPC	NUIDUCD-CeADAR, ATOS IT, TUBS, FDI, ARCADa, KU Leuven, LAU)
5	ATOS IT	Electric power and energy consumption metric consumed on bare metal hosts	T5.2	ATOS IT	NUIDUCD-CeADAR, UPC, TUBS, FDI, Fraunhofer, INRIA, PAL ROBOTICS,

WP	WPL	Data	Tasks	WTL	Responsible Partners
					Bit&Brain, ARX.NET S.A. and UPSaclay
		Templates and Self-assessments	T5.3	FDI	NUIDUCD-CeADAR, UPC, ATOS IT, EVIDEN RO, NCSR "D", INRIA, ARCADA, KU Leuven, LAU, PAL ROBOTICS, Bit&Brain, ARX.NET S.A.
6	PAL ROBOTICS	Conversations	T6.2	Bit&Brain	NUIDUCD-CeADAR, UPC, ATOS IT, TUBS, NCSR "D", FDI, Fraunhofer, INRIA, ARCADA, KU Leuven, LAU, PAL ROBOTICS, ARX.NET S.A., Q-PLAN
		Robot state information			
		3D information of the environment			
		2D information of the environment			
		Model's performance in different nodes of the topology			
		Image classification, photograph tagging, and visual cues for predictions			
		Face detection, face retrieval			
		Sleep recordings for sleep stage decoding			
		Sleep recordings during real-time closed-loop auditory stimulation			
		Use cases data evaluation and validation	T6.3	LAU	All partners
7	Q-PLAN	Website and social media analytics	T7.1	Q-PLAN	All partners
		Data collected from project events	T7.1, T7.4	Q-PLAN	All partners
		Data collected from dissemination and communication actions	T7.1	Q-PLAN	All partners
		Newsletter subscriptions	T7.1	Q-PLAN	All partners
		Market analysis, Competition analysis, as well as financial planning	T7.3	Q-PLAN	NUIDUCD-CeADAR, ATOS IT, FDI, Fraunhofer, PAL ROBOTICS, Bit&Brain, ARX.NET S.A., EIT DIGITAL
		Constituting the EIAB	T8.2	KU Leuven	All partners and UPSaclay
		Management of the EIAB			

6. Data security

MANOLO securely handles any collected/generated or reused data throughout its entire lifecycle, as it is essential to safeguard this data against accidental loss and/or unauthorised access. To achieve this, the project applies appropriate technical and organisational measures based on a risk assessment of the relevant data that considers the impact and the likelihood of a potential data breach. With that in mind, the project's data security strategy aims at minimizing the probability that a data breach will occur during the course and after the completion of MANOLO, either from human error or hardware failure, as well as inhibiting any unauthorised access. Particularly, in the case of personal data collection/generation, it is crucial that this **data can only be accessed by those authorised to do so**.

All project partners are responsible for processing³⁸ data using appropriate means, such as private servers or cloud service providers that adhere to the relevant legal data protection requirements (e.g., GDPR) and ensure that these **data are protected**, and any **necessary data security controls have been implemented**, to minimize the risk of information leak and destruction (including ARX.NET's cloud repository). This case refers to the data that are closed and therefore are not shared and/or reused within the framework of the project. In this case, to minimize the consequences of potential data losses, the data are **backed up at regular time intervals based on change frequency and criticality. The backed-up files are stored in appropriate storage media, including external hard drives, flash drives, NAS devices, and reputable cloud services**, to safeguard their preservation, while also enabling their recovery at any time. Moreover, **integrity checks**³⁹ are carried out regularly, ensuring that the stored data have not been changed or corrupted.

Access to closed data is only permitted to authorised project partners. In case there is a personal data breach, the responsible project partner will notify, without undue delay and, where feasible, not later than 72 hours after having become aware of it, its competent national supervisory authority (e.g., data protection authority) as well as the data subject(s) that may be affected by the breach. Moreover, the responsible partner will document any personal data breaches, including information such as the facts relevant to the breach, its effects, and the remedial action(s) taken.

Identification and authentication access controls play an important role in the context of the project, as they help partners protect the data collected/generated during MANOLO, especially personal data. To this end, each project partner is responsible for and committed to applying appropriate access controls to the data they are processing. Finally, to safeguard the privacy of the users of the MANOLO website, a dedicated **privacy policy** defines how this online space collects, processes, and uses personal data, the security procedures followed, the users' rights, as well as the cookies policy employed.

On another note, openly available data are stored safely for long-term preservation on Zenodo, in the same cloud infrastructure as research data from CERN's Large Hadron Collider, using CERN's battle-tested repository software INVENIO, which is used by some of the world's largest repositories (such as INSPIRE HEP and the CERN Document Server). Along these lines, data are stored and backed up in CERN's EOS service in an 18-petabyte disk cluster. Both data files and metadata are kept in multiple online replicas and independent replicas, ensuring their long-term preservation as well as their recovery when necessary. Moreover, for each file, two independent MD5 checksums are stored. One checksum is stored by INVENIO, used to detect changes to files made from outside of it, whereas the other checksum is stored by EOS, and used for automatic

³⁸ Processing, according to Regulation (EU) 2016/679 of the European Parliament (General Data Protection Regulation), means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

³⁹ An integrity check is the process of comparing the current state of stored data and/or programs to a previously recorded state in order to detect any changes.

detection and recovery of file corruption on disks. In this context, access control is applied by the different levels of openness that Zenodo allows (i.e., open, embargoed, restricted, and closed).

Accordingly, data stored at GitHub are also stored safely for long-term preservation. In general, GitHub is a GDPR-compliant platform that ensures data privacy and security. GitHub is a Trusted Cloud Provider (TM) with the Cloud Security Alliance (CSA)⁴⁰, whereas its Information Security Management System (ISMS) has been certified against the ISO/IEC 27001:2013 standard.

⁴⁰ <https://cloudsecurityalliance.org/>

7. Ethics and other issues

This Chapter addresses the ethical aspects of MANOLO's Data Management Plan and the ethical compliance of the underlying data foreseen to be collected/generated or reused under the project's activities. The project processes data that are not included in any special category of personal data (i.e., non-sensitive data) according to the relevant data protection legislation (e.g., GDPR). Under **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (GDPR)**, all personal data processed for the project's activities shall be:

- processed lawfully, fairly, and in a transparent manner about the data subject;
- collected for specified, explicit, and legitimate purposes relative to the project's objectives and not further processed in a manner that is incompatible with those purposes;
- adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed;
- accurate and, where necessary, kept up to date;
- kept in a form that permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed;
- processed in a manner that ensures appropriate security of the personal data (see section 6).

For all personal data processing activities within the framework of the project, at least one lawful basis as of Art. 6 GDPR applies. Where informed consent is chosen as the lawful basis for processing, all relevant provisions of the data protection legislation (e.g., Art.7 GDPR) are observed. Moreover, **personal data processing carried out by partners is in line with relevant EU and national regulations**. The project's Privacy Policy and the templates of the Informed Consent Form and the Data Subject Request Form, used in the implementation of the project's activities, are compliant with the General Data Protection Regulation and annexed to this DMP (see Annex). Last but not least, **no transfer of personal data outside the EU is foreseen as part of the project's implementation**. In the case of data storage providers situated both inside and outside the EEA, partners are committed to ensuring their compliance with the relevant GDPR requirements before starting to use their services.

It is important to highlight that each partner is responsible for ensuring that the templates for the Informed Consent Form and Subject Data Request Form (including references to the project's Privacy Policy and any other applicable specific privacy policies) are appropriately adjusted according to (i) the needs of the activity for which they are being used by them as well as to (ii) the relevant data protection laws and regulations applicable to their respective countries and/or organisation. All partners should keep records to demonstrate that data subjects have consented to the processing of their personal data and use consent management mechanisms that make it easy for individuals to withdraw their consent.

Finally, **no other national/funder/sectoral/departmental procedures for data management were used in the framework of MANOLO**.

8. Conclusions and way forward

This interim version of the MANOLO DMP aims at safeguarding the sound management of the data collected, processed, and/or generated during the project's activities across their entire lifecycle, while also making them FAIR. It describes all the underlying processes of the MANOLO data management, collection, process, and generation, following the GDPR guidelines, and sheds light on (i) the data being collected, processed, and/or generated under the project activities, (ii) the specific objectives under which each dataset is collected, processed and/or generated, (iii) the management of the other research outputs of the project (iv) the allocation of resources and data management responsibilities and (v) the data security and ethical aspects of the data.

In the framework of MANOLO, the DMP is a living document and is updated throughout the project, considering its latest developments and available results. It is expected to be further developed and updated at least once by the end of the project (i.e., as D8.4 in M36). If necessary, additional ad hoc updates may be realised to include new data, better detail, and/or reflect modifications in the methodologies applied or other aspects relevant to data management (such as costs for making data FAIR, size of data, etc.), changes in consortium policies and plans or other potential external factors. In the upcoming version, we also plan to incorporate additional synthetic datasets originating from public/open resources or the use cases to further enhance the comprehensiveness and applicability of the information provided.

Annexes

Annex I – Privacy Policy

Within this Annex, the reader can find the draft project's overall Privacy Policy, which is uploaded on the project's website. Task leaders are responsible for developing any additional privacy policy needed in their tasks and activities in close collaboration with the PC. Moreover, they are responsible for appropriately adjusting and translating the templates to fit the needs and specificities of their task's activities.

PRIVACY POLICY

LAST UPDATE: < date of the privacy policy's last update in the form of DD/MM/YYYY >

1. Who we are

MANOLO, a Horizon Research & Innovation Action project, is funded by the European Union, under Grant Agreement No. 101135782, with a budget of approximately €8.5 million. The project started in January 2024 and will last 36 months.

MANOLO will deliver a complete stack of trustworthy algorithms and tools to help AI systems reach better efficiency and seamless optimization in their operations, resources and data required to train, deploy and run high-quality and lighter AI models in both centralised and cloud-edge distributed environments. It will push the state of the art in the development of a collection of complementary algorithms for training, understanding, compressing and optimising machine learning models by advancing research in the areas of: model compression, meta-learning (few-shot learning), domain adaptation, frugal neural network search and growth and neuromorphic models. Novel dynamic algorithms for data/energy efficient and policy-compliance allocation of AI tasks to assets and resources in the cloud-edge continuum will be designed, allowing for trustworthy widespread deployment. To support these activities, a data management framework for distributed tracking of assets and their provenance (data, models, algorithms) and a benchmark system to monitor, evaluate and compare new AI algorithms and model deployments will be developed. Trustworthiness evaluation mechanisms will be embedded at its core for explainability, robustness and security of models while using the Z-Inspection methodology for Trustworthy AI assessment, helping AI systems conform to the new AI Act regulation. MANOLO will be deployed as a toolset and tested in lab environments via Use Cases with different distributed AI paradigms within cloud-edge continuum settings; it will be validated in verticals such as health, manufacturing, and telecommunications aligned with ADRA identified market opportunities, and with a granular set of embedded devices covering robotics, smartphones, IoT as well as using Neuromorphic chips. MANOLO will integrate with ongoing projects at EU level developing the next operating system for cloud-edge continuum while promoting its sustainability via the AI-on-demand platform and EU portals.

The project partners of the MANOLO consortium, listed below, process certain types of personal data for the purposes of the project. Each partner is responsible for the personal data they collect and process during their activities under the framework of the project:

- University college Dublin, National University of Ireland (NUIDUCD-CeADAR) www.ceadar.ie, Ireland
- Universitat Politècnica de Catalunya (UPC) <https://www.upc.edu>, Spain
- ATOS IT Solutions and Services Iberia SL (ATOS IT), <https://atos.net/es/spai>, Spain

- Eviden Technologies Srl (EVIDEN RO) <https://eviden.com/>, Spain
- Technische Universitaet Braunschweig (TUBS) <https://www.tu-braunschweig.de/en/>, Germany
- National Center for scientific research “DEMOKRITOS” (NCSR “D”) <https://www.iit.demokritos.gr/>, Greece
- Four Dot Infinity information and telecommunications solutions private company (FDI) <https://fourdotinfinity.com/>, Greece
- Fraunhofer Gesellschaft Zur Forederung der Angewandted Forschung EVRAU (Fraunhofer) <https://www.iis.fraunhofer.de/en.html>, Germany
- National Institute for Research in Digital Science and Technology (INRIA) <https://www.inria.fr/en>, France
- UPSaclay Universite Paris-Saclay (UPSaclay Universite Paris-Saclay), <https://www.universite-paris-saclay.fr/en>, France
- Yrkeshogskolan Arcada AB (ARCADA) <https://www.arcada.fi/en>, Finland
- Katholieke Universiteit Leuven (KU Leuven) <https://www.kuleuven.be/english/kuleuven>, Belgium
- Laurea-Ammattikorkeakoulu OY (LAU) <https://www.laurea.fi/en/>, Finland
- Pal Robotics PL (PAL ROBOTICS) <https://pal-robotics.com/>, Spain
- BIT & BRAIN Technologies SL (BIT&BRAIN) <https://www.bitbrain.com/>, Spain
- ARX NET AE Ypiresies kai epichirisis diadiktyou anonimi etairia (ARX.NET S.A.) <https://web.arx.net/>, Greece
- Q-PLAN International Advisors PC (Q-PLAN) <https://qplan-intl.gr/>, Greece
- EIT DIGITAL (EIT DIGITAL) <https://www.eitdigital.eu/>, Belgium

For further information, we can be contacted at: manolo-all@listserv.heanet.ie.

2. How we collect your personal data

We collect personal data both directly and indirectly:

- We obtain personal data directly from individuals in a variety of ways, including the following cases:
 - an individual subscribes to our newsletter/s;
 - an individual registers to attend meetings, workshops and other events we host and during attendance at such events;
 - we establish cooperative relationships with an individual;
 - we provide professional services pursuant to our contract with the European Commission;
- We obtain personal data indirectly about individuals from a variety of sources, including:
 - our research partners;
 - our networks and contacts;
 - public and open data sources such as public registers, news articles and internet searches;
 - social and professional networking sites (e.g., LinkedIn).

3. What types of data we collect

We only collect the data necessary for our project’s smooth implementation. These data fall into the following categories:

- contact details (name/ surname, e-mail address, phone number);
- professional information (job title, organisation, field of expertise);
- demographics (e.g. age, gender, nationality);
- information about what a person knows or believes;

- videos and photos (from people who attend our events, webinars, workshops, etc.).

4. Bases of lawful processing

We process personal data on the following legal bases:

- Legal obligations – for processing activities required for compliance with both applicable national and European legislation, as well as with the specific legal and regulatory framework of the Horizon Europe Framework Programme for Research and Innovation of the European Union.
- Consent – for processing activities such as the organisation of meetings, workshops, other events, and the dissemination of the project's results.
- Contractual obligations – for processing activities such as reporting to the European Commission and complying with the project's publicity obligations.

5. What we do with your personal data

We process your personal data with the purpose of:

- Disseminating our project's results to different types of stakeholders;
- Sending invitations and providing access to guests attending our meetings, events, and workshops;
- Administering, maintaining, and ensuring the security of our information systems, applications, and websites;
- Processing online requests or queries, including responding to communications from individuals;
- Complying with contractual, legal, and regulatory obligations.

6. How we secure your personal data when we process it

We continuously apply a personal data risk assessment process to identify, analyse, and evaluate the security risks that may threaten your personal data. Based on the results of this risk assessment, we define and apply a set of both technical and organizational measures to mitigate the above security risks, including but not limited to:

- Data Protection Policies to guide personnel involved in the MANOLO project when processing your data;
- Written contracts with organisations that process personal data on our behalf;
- Non-Disclosure Agreements with personnel of the MANOLO consortium partners;
- Backup process, antimalware protection, access control mechanisms, etc.
- Some of MANOLO's partners have appointed a Data Protection Officer.

7. Do we share personal data with third parties?

Your personal data may be shared with trusted third parties to help us deliver efficient and quality services. When we do so, we ensure that recipients are contractually bound to safeguard the data we entrust to them before we share the data. We may engage with several or all of the following categories of recipients:

- Parties that support us as we provide our services (e.g., cloud-based software services such as Dropbox, Microsoft SharePoint, Google);
- Our professional advisers, including lawyers, auditors, and insurers;
- Dissemination service providers (e.g., MailChimp);

- Law enforcement or other government and regulatory agencies or other third parties as required by, and in accordance with, applicable law or regulation;
- The European Commission, according to our relevant contractual obligations.

8. Do we transfer your personal data outside the European Economic Area?

We do not own file servers located outside the European Economic Area (EEA). However, some partners may use cloud and/or marketing services from reputable providers such as SharePoint, Dropbox, MailChimp, Google, etc., situated both inside and outside the EEA. We always check that such providers comply with the relevant GDPR requirements before starting to use their services.

9. Do we use cookies?

Our website, <https://www.manoloproject.eu>, uses cookies. We use cookies to improve site navigation and to analyse our traffic. In the next paragraphs, you will learn what cookies are, what cookies we use and how you can disable their use.

Cookies are small alphanumeric characters files that are transferred from websites that you visit to your PC, smartphone, tablet or any other electronic device that you use for browsing. Cookies assign your computer with a unique ID, which in turn becomes your own identity each time you return to a website. Cookies do not damage your electronic equipment, nor can they read information from other files on your computer's hard drive.

Websites use cookies to “remember” for a period your actions and preferences such as language, font size and other display features. In this way, you are not required to enter these preferences each time you visit a website and navigate its pages. Furthermore, cookies help website owners to analyse how you are using their website and whether or not you are facing problems as you navigate them.

Personal Data Protection legislation states that we can store cookies on your device if they are strictly necessary for the operation of our website. For all other types of cookies, we need your consent. You can change or withdraw your consent at any time from the Cookies Settings on our website.

On our website, <https://www.manoloproject.eu>, we use necessary, statistics and functional cookies. Some cookies are placed by third-party services (e.g., Google Analytics).

Without the necessary cookies, the website cannot perform its basic functions. These cookies are generally placed automatically either when a webpage is loaded or as a result of a user's request (that cannot be fulfilled without using a relevant cookie). Usually, necessary cookies expire when you close your browser.

Statistics cookies help us to measure our website's traffic and where it comes from so that we can improve its performance. Moreover, they collect information about which pages are more popular and how visitors use our website. None of this information can be used to identify you. It is all aggregated and, therefore, anonymized. This category includes cookies from third-party analytics services such as Google Analytics. Our website uses the following statistics cookies:

Name	Provider	Purpose	Expiry
_ga	Google Analytics	This cookie is used to distinguish unique users by assigning a randomly generated number as a client identifier. It is included in each page request on a site and used to calculate visitor, session and campaign data for the site's analytics reports.	2 years

Name	Provider	Purpose	Expiry
_gid	Google Analytics	This cookie stores and updates a unique value for each page visited. It is used to generate statistical data on how the visitor uses the website. It collects anonymized data including the number and origin of visitors and the websites they visited.	1 day
_gat	Google Analytics	This cookie is used to throttle the request rate, that is limiting the collection of data on high-traffic sites.	1 minute

Functionality cookies allow websites to remember the user's site preferences and choices they make on the site, including username, region, and language. This allows the website to provide personalized features if you share your location. They are anonymous and do not track browsing activity across other websites.

You can delete cookies and stop them from being installed in your browser. There is no standardized way to remove cookies since different browsers clear cookies using different procedures. Please follow the instructions provided by each browser manufacturer on how to remove cookies.

If you wish to stop sending your information to Google Analytics, you may install in your browser the Google Analytics Opt-out Browser Add-on.

You may select and change the cookie settings (with the exception of the technically necessary cookies) by clicking on Cookie settings. By making a cookie active, you consent to its use according to the provisions of this Cookies Policy.

10. Your rights

You have the following rights regarding our processing of your personal data:

- Right to withdraw consent – You can withdraw consent that you have previously given to one or more specified purposes to process your personal data. This will not affect the lawfulness of any processing carried out before you withdraw your consent.
- Right of access – You can ask us to verify whether we are processing personal data about you and if so, to have access to a copy of such data.
- Right to rectification and erasure – You can ask us to correct our records if you believe they contain incorrect or incomplete information about you or ask us to erase your personal data after you withdraw your consent to processing or when we no longer need it for the purpose it was originally collected.
- Right to restriction of processing – You can ask us to temporarily restrict our processing of your personal data if you contest the accuracy of your personal data, prefer to restrict its use rather than having us erase it, or need us to preserve it for you to establish, exercise or defend a legal claim. A temporary restriction may apply while verifying whether we have overriding legitimate grounds to process it. You can ask us to inform you before we lift that temporary processing restriction.
- Right to data portability – In some circumstances, where you have provided personal data to us, you can ask us to transmit that personal data (in a structured, commonly used, and machine-readable format) directly to another entity.
- Right to object – You can object to our use of your personal data for direct marketing purposes, including profiling or where processing has taken the form of automated decision-making. However, we may need to keep some minimal information (e.g., e-mail address) to comply with your request to cease marketing to you.

- Right to make a complaint to your local Data Protection Authority (DPA) (see https://ec.europa.eu/justice/article-29/structure/data-protection-authorities/index_en.htm) regarding any concerns you may have about our data handling practices.

To ask us to do anything of the above, you can contact us by email: info@manoloproject.eu. We will promptly examine your request against the relevant requirements of the laws and regulations governing privacy and personal data protection and we will answer the latest within 30 days after receiving your request. We will ask you for some kind of identification (e.g. photocopy of your identity card or passport) to avoid the non-authorized reveal of your personal data. If, for reasons of the complexity of the request or a multitude of requests, we are unable to respond promptly, we will notify you within 30 days of any delay, which in no case may exceed two months from the expiration of the 30-day deadline.

11. How long do we retain personal data?

We retain personal data to provide our services, stay in contact with you and comply with applicable laws, regulations, and contractual obligations to which we are subject. Please note that we have an obligation to retain data concerning projects funded by the Horizon Europe Framework Programme for Research and Innovation of the European Union for up to five years after the end of the project (unless auditors request further retention). After the expiry of the retention period, and unless further legitimate grounds for retention arise, we will dispose of personal data in a secure manner.

12. Disclaimer of liability for third-party websites

Although our site may contain links to third-party sites, including the sites of the MANOLO consortium partners, we are not responsible for the privacy practices or content of these sites, and we expressly disclaim any liability for any loss or damage that may be caused by the use of these links. We do not monitor the privacy practices or the content of these sites. If you have any questions about the privacy practices of another site, you should contact the site's responsible personnel. We suggest you read the privacy policy of each website you interact with, before allowing the collection and use of your personal data.

We may also provide social media features that allow you to share information on your social networks and interact with the MANOLO project on various social media sites. The use of these social media features may result in the collection or sharing of information about you. We recommend that you check the privacy policies and regulations of the social networking sites you interact with to ensure that you understand what information may be collected, used and disclosed by these sites.

13. Children

We do not knowingly collect, use, or disclose information from children under the age of 16. If we learn that we have collected the personal information of a child under 16 we will take steps to delete the information as soon as possible. Please immediately contact us if you become aware that a child under 16 has provided us with personal information.

14. Revisions of this Privacy Policy

This Privacy Policy is valid from 1/1/2024 and replaces any other previous notifications that we had issued in the past regarding our personal data management practices. We reserve the right to revise this Policy at any time. The current version will be always uploaded to our website indicating the date of entry into force, so you know when the most recent revision took place. If there are critical changes in this policy or our personal data practices change significantly in the future, we will notify you by posting the changes on our website.

Annex II – Informed Consent Form

Text in red colour contains guidelines for adjusting this template and should be deleted.

Text included in < > and/or highlighted with yellow should be replaced with content that is suitable to the context of each activity & project as well as to the organisation seeking to obtain the consent.

Before using this template take the time to carefully read and adjust it to the needs of the activity at hand as well as to any relevant regulations and particularities applicable to your country and organisation. Specifically, this template has been developed for a data processing activity involving interviews, however you can easily adapt it to other common data processing activities such as surveys, events organizing, etc.

INFORMED CONSENT FORM

Who we are:

We are < Insert Partner Name > and we are contacting you in the framework of MANOLO, a project funded by the European Union under the Horizon Europe Research and Innovation funding programme. A detailed description on how MANOLO handles personal data is presented in the project's Privacy Policy that accompanies this Consent Form.

Project: MANOLO - Trustworthy Efficient AI for Cloud-Edge Computing (GA Number 101135782).

Partner:

Organisation name: < Insert Partner Name >

Address: < Insert Partner Address >.

Phone: < Insert Partner Phone >.

E-mail: < Insert Partner Generic E-mail Address >

Responsible persons:

You may delete the line referring to the Data Protection Officer if your organisation does not have one.

#	Role	Name	E-mail
1	MANOLO Project Manager	<Insert name of project manager>	<Insert email of project manager>
2	Interviewer	<Insert name of interviewer>	<Insert e-mail of interviewer>
3	Data Protection Officer	<Insert name of DPO >	<Insert e-mail of DPO >

What do we need from you?

Please explain in a brief paragraph (4-5 lines) the activity and its purpose under the frame of the project. You can find below an indicative example.

Example: We need you to participate in an interview that will be carried out by MANOLO with a view to monitor and assess MANOLO' outcomes, impacts and perceptions change towards a trustworthy Efficient AI for Cloud-Edge Computing.

The interview is expected to last for no more than < Insert number of minutes > minutes. We will take written notes and we will be making a sound recording of the interview.

Please adapt the following text to accurately depict the type of personal data to be collected.

To effectively conduct this interview, we need to process some of your personal data:

- Your contact details (full name, email, phone number);
- Some basic demographics (age, gender);
- Your professional info (organization, job position, field of expertise);
- Your opinions on the subject matter.

Why do we need your data & what will we do with them?

We need your data to contact you in order to plan and carry out the aforementioned interview and to resolve any ambiguities, questions and other issues that may arise after and as a result of the interview. We also need to record your data to keep track of the interview process. The project's deliverables that will be derived by the interview will not include your personal data or any other information that could identify you. Your personal data will remain on our written notes (interview's transcript) and/or the sound recording we will make during the interview.

We will share your data with a few other MANOLO project partners that are also involved in this task and will participate in the drafting of the relevant deliverables. We are also obliged to grant access to your data to:

- EU officials such as our Project Officer for purposes related to the project's evaluation;
- EU agencies and other authorities for project's auditing purposes.

We would also be very happy if you gave us your consent to contact you in the future to ask you to participate in other project's activities (e.g., surveys, interviews, project events, MANOLO Industrial Advisory Board, etc.) and also to inform you about the project's progress (e.g., by sending you a newsletter or similar messages).

How can you withdraw your consent?

You should know that you can withdraw your consent at any time by communicating either on the phone or by email with the responsible persons listed on the previous page. With regards to the informational messages and newsletters, you can always opt out by simply clicking the link "Unsubscribe" or something similar included at the end of all the relevant messages.

I hereby give my consent to the processing of my personal data needed for:

*(Please, tick the boxes below to confirm that you give us your consent for the respective subject. Any boxes left unticked mean that **you do not consent to the relevant subject.**)*

#	Consent Subject	Tick box
1	My participation in an interview that will be carried out by MANOLO to < insert key objective of the interview >	
2	My participation in future activities of MANOLO	
3	Receiving newsletters and messages regarding MANOLO activities	

Name of participant
Signature

Date

Annex III – Data Subject Request Form

Text in red colour contains guidelines for adjusting this template and should be deleted.

Text included in < > and/or highlighted with yellow should be replaced with content that is suitable to the context of each activity & project as well as to the organisation seeking to obtain the consent.

MANOLO

Data Subject Request form

You may delete the data referring to the Data Protection Officer if your organisation does not have one.

Contact

<Insert name of responsible Project Manager>	<Insert name of DPO > (Data Protection Officer)
<Insert email of responsible Project Manager>	<Insert e-mail of DPO >

Data Subject Request Form

This form should be used to submit a data subject request under the provisions of the European Union General Data Protection Regulation (GDPR).

Submitter Details

Title:	
Name:	
Address:	

Type of Request

Please select the type of request you are making:

☐	<i>Consent Withdrawal</i>
☐	<i>Access request</i>
☐	<i>Rectification of personal data</i>
☐	<i>Erasure of personal data</i>
☐	<i>Restriction of processing of personal data</i>
☐	<i>Personal data portability request</i>
☐	<i>Objection to processing of personal data</i>
☐	<i>Request regarding automated decision making and profiling</i>

Personal data involved

--

Request details

Request reason/justification

Name:

Signature:

Date:

Once completed, this form should be submitted via e-mail to < Insert contact e-mail of Partner > or posted to:

< Insert Partner Name >

< Insert Partner Address >

Annex IV- Main changes in the data collected/generated since M4

The following table summarizes the Main changes that have been made regarding the MANOLO datasets. Beyond the following changes, several minor updates to better describe the datasets as well as reflect the type, format, size, availability and accessibility were made.

Name of activity	Data	Status
Identification of insights about the needs, available or forthcoming technological solutions, key actors, and openly available resources	Desk research and interviews	Updated
	Survey's results	Updated
	Data from the State of the Art and regulatory resources	Updated
Data Distillation & Synthetic Feature Representation	Synthetic/Distilled dataset	Updated
Model Pruning, Quantization, and Knowledge Distillation	LLM dataset and its generation results	New
Domain Adaptive Few-Shot Continuous Learning	Time series with the evolution of metrics and statistics	Deleted
	T-Hyperparameter search on compression algorithms, with results and dataset + model metadata	New
	Curated texts for LLMs	New
	Generated responses from LLMs	New
AI/ML Models and functions allocation on the cloud-edge continuum	Deployment information after optimisation	Updated
Develop methods and tools for human oversight in monitoring and validating AI systems within the cloud-edge continuum	Templates and Self-assessments	Updated
Interaction with humans for assisted living	Model's performance in different nodes of the topology	New
Mobile computer vision; classification	Image classification, photograph tagging, and visual cues for predictions	Updated
Mobile computer vision; Face detection	Face detection, face retrieval	New
Auditory stimulation for memory consolidation	Sleep recordings for sleep stage decoding	Updated
	Sleep recordings during real-time closed-loop auditory stimulation	Updated
Evaluation and Validation of MANOLO Solution	Use cases, data evaluation, and validation	Updated



MANOLO

CLOUD-EDGE EFFICIENT & TRUSTWORTHY AI

GA 101135782

Partners

