Deliverable D1.19: Data Management Plan (DMP)

WP 1 – Project management and coordination

Authors: Zoi Arkouli (LMS), Kosmas Alexopoulos (LMS)

Co-funded by the European Union

Grant Agreement number: 101100707

Disclaimer

Document Information

G.A. No.	101100707	Acronym	AI-Matters
Full Title	AI MAnufacturing Testing and experimenTation network For EuRopean industrieS		
Call	DIGITAL-2022-CLOUD-AI-02		
Type of Action	DIGITAL Simple Grants		
Start Date	1st Jan 2023	**Duration**	60 months
Project URL	-		
Document URL	-		
EU Project Officer	Mariusz Bałdyga, Programme Officer, EU policies, DG CONNECT, Unit A1		
Project Coordinator	Valentina Ivanova, CEA-LIST, Deputy Director in charge of European affairs		
Deliverable	Deliverable D1.19: Data Management Plan (DMP)		
Work Package	WP 1 – Project management and coordination		
Date of Delivery	**Contractual**	M6	**Actual** M8
Nature	R - Report	**Dissemination Level**	PU -Public
Lead Beneficiary	Laboratory for Manufacturing Systems and Automation (LMS)		
Lead Author	Zoi Arkouli	**Email**	Arkouli@lms.mech.upatras.gr
	LMS	**Phone**	+30-2610-910160
Other authors	Kosmas Alexopoulos (LMS)		
Reviewer(s)	Rosamaria Russo (Engineering), Farzam Ranjbaran (CEA), David Servat (CEA)		
Keywords	data management, data protection, data privacy, data security		

Document History

Version	Issue Date	Stage	Comments	Contributor
0.1	17.05	Draft	Document outline	LMS
0.2	24.05	Draft	Document executive summary, introduction	LMS
0.3	22.06	Draft	Full draft ready	LMS
0.4	23.06	Under review	Comments have been added where partners need to confirm/review the except of interest	LMS, ENGINEERING, CEA
1.0	30.06	Published		

Table of Contents

Executive summary 4
Glossary 5
1 Introduction 6
1.1. About AI-Matters 6
1.2. Mission & Vision 6
1.3. Purpose of the deliverable 6
2 DATA SUMMARY 8
2.1 Purpose of data collection/generation 9
2.2 Data sources 11
2.2.1 Types and formats of the collected/generated data 11
2.2.1.1 Service Request Identification 12
2.2.1.2 Dataset Overview 12
2.2.1.3 Dataset Format 12
2.2.1.4 FAIR Dataset 13
2.2.1.5 Dataset Security 13
2.2.2 Collected dataset 13
2.3 Relation of project objectives and data collection 13
2.4 Data security 14
2.4.1 Reuse of existing data 15
2.5 Data utility 15
2.6 Intellectual property rights 16
2.7 Research data and Open Access 16
3 FAIR DATA 17
3.1 Data discoverability 17
3.1.1 Data identification mechanisms 18
3.1.2 Naming conventions 18
3.1.3 Keywords 20
3.1.4 Document versioning 21
3.1.5 Metadata standards 22
3.2 Data accessibility 23
3.2.1 Openly available data 23
3.2.2 Procedures to access data 23
3.2.3 Handling access to restricted data 23

3.2.4 Data repository 23
3.3 Data interoperability 24
3.3.1 Ascertainment of data interoperability 24
3.3.2 Vocabulary used 24
3.4 Enhancement of data reuse 25
3.4.1 Data use by third parties 25
3.4.2 Data availability for reuse 26
3.4.3 Data quality assurance 26
4 RESOURCE ALLOCATION 27
4.1 Data management responsibility and related costs 27
4.2 Costs estimations for long-term data preservation 27
5 SECURITY 27
6 ETHICS 28
6.1 Data related ethics screening aspects 28
6.2 Personal data processing 29
6.2.1 Anonymisation/pseudonymisation techniques 30
6.2.2 KPIs & Personal data protection 30
7 Conclusions 31
8 Annexes 32
8.1 Annex I: Information sheet 32
8.2 Annex II: Consent form 34
8.3 Annex III: Dataset Template 35
8.4 Annex IV: Disclaimers for Dissemination & Communication activities 38

Table of Figures

Figure 1: AI-MATTERS Hierarchical Project Objectives (overarching, High-level and Project Objectives) 14

Table of Tables

Table 1 – D1.19 glossary 5
Table 2. Document template extract showing the history panel 21
Table 3. Document template extract showing the history panel 21

Executive summary

This deliverable aims to provide a single point of reference on all aspects related to data management pertaining to the activities of the AI-MATTERS project. Given that the AI-MATTERS network is structured into seven nodes and one satellite, where each represents a national level and is a cluster of consortium parties responsible to support the deployment of AI solutions in the assigned jurisdiction, the data management of data relevant to the services provision will be performed at the nodes and satellite level. The coordinator will be responsible to manage data related to the website of the project, as well as data associated with project management given that the relevant platforms are provided by CEA. Therefore, the current deliverable (D1.19) suggests guidelines for data management to be used by aforementioned entities. In particular, D1.19 outlines how the data should be handled during and after the project lifetime presenting the principles on which data management will be based, as well as the procedures, measures, and results to ensure that data management is compliant to the relevant local, EU, and international standards. Moreover, the consortium will strive to make the research data as open as possible and confidential as necessary, while compliant with the FAIR principles. Due to the delivery of the DMP at the early stage of the project, the first version of the data management plan is founded on the expectations of the consortium based on the initially defined services in the context of AI-MATTERS. The document will be updated as per the needs that arise during the project's lifespan, whereas any updates that occur will be recorded in the project management reports.

Glossary

For the purposes of this document, the following terms and definitions apply.

1	AI:	Artificial Intelligence
2	CA:	Consortium Agreement
3	CC:	Creative Commons
4	DMP:	Data Management Plan
5	DoA:	Description of Action
6	DOI:	Digital Object Identifier
7	DPO:	Data Protection Officer
8	EA:	Ethics Advisor
9	ESCC	Ethics Screening and Compliance Committee
10	EthSR:	Ethics Summary Report
11	EU:	European Union
12	FAIR:	Findable, Accessible, Interoperable and Re-usable
13	GA:	Grant Agreement
14	GDPR:	General Data Protection Regulation
15	HPC:	High Performance Computing
16	HRC:	Human Robot Collaboration
17	IPR:	Intellectual Property Rights
18	IT:	Information Technology
19	KPI:	Key performance indicator
20	NDA:	Non Disclosure Agreement
21	PII	Personally Identifiable Information
22	OSH:	Occupational Safety and Health
23	QA:	Quality Assurance
24	QC:	Quality Control
25	SMEs:	Small and medium-sized enterprises
26	TEF:	Testing and Experimentation Facilities
27	URL:	Uniform Resource Locator
28	WP:	Work Package

Table 1 – D1.19 glossary

1 Introduction

1.1. About AI-Matters

Framed under the DIGITAL Europe Programme, the project AI in Manufacturing TesTing and experimentation facilities for EuROpean SMEs (AI-Matters) is a collaboration between 25 consortium partners from eight European countries (Germany, Italy, The Netherlands, Czech Republic, Greece, Spain and France), forming seven nodes and one satellite. It aims at increasing the resilience and the flexibility of the European manufacturing sector through the deployment of the latest developments in AI and robotics, and intelligent, autonomous systems for flexible production. To this end, AI-Matters will offer its customers an extensive service catalogue spanning the topics above that evolves through continuous updates as needs and expectations of the European manufacturing industry progress.

1.2. Mission & Vision

The vision of the AI-Matters project under the DIGITAL Europe Programme is to enhance the leadership of the European manufacturing industry through the uptake of innovation.

The overarching mission of the AI-MATTERS network is to implement concrete steps to strengthen Europe's positioning and leadership in the uptake of human-centric Artificial Intelligence systems. enabling robotics and other advanced technologies to create leap transformations of the EU manufacturing system.

More specifically, AI-MATTERS sets out:

- To create a world-class network of large-scale reference sites for testing and experimentation of AI powered solutions
- To enable pathways to impact for the deployment of trustworthy, transferable and scalable Industrial AI in Europe
- To accelerate the transition towards a more AI-powered manufacturing industry and improve the quality and sustainability of production
- To advance technologies and capabilities in and across three main topics: Factory-level optimization, Collaborative robotics, Circular economy, but also in other impactful areas where adoption of AI-enabled technologies in manufacturing need boosting
- To establish/enrich collaborative links/synergies with other EC-funded networks and programmes

1.3. Purpose of the deliverable

In the context of the AI-MATTERS activities, all partners of the consortium should adhere to sound data management in order to ensure the meaningful data collection, processing and/or generation throughout the duration of the project, as well as that the relevant procedures are well-managed, archived and preserved, in line with the Guidelines on Data Management in Horizon 2020 and according to the EU General Data Protection Regulation (GDPR). Furthermore, the consortium will strive to make the research data as

open as possible and confidential as necessary, while compliant with the FAIR principles Along these lines, D1.19 aims to achieve the following objectives:

- Describe the data management lifecycle for the data to be collected and/or generated in the framework of AI-MATTERS, serving as a key guideline for data management.
- Outline the methodology to safeguard the management of the data collected, and/or generated and to handle the data under the FAIR principles.
- Provide information on the data that will be collected and/or generated together with the manner based on which data will be handled during and after the end of the project along with the standards applied to this end.
- Describe what kind of data and the way for making them accessible and searchable to interested stakeholders and their curation and preservation.
- Present an estimation of the resources allocated to make data FAIR, while also identifying the responsibilities pertaining to data management and addressing data security.
- Provide guardrails on ethical aspects that relate to data management
- Provide templates of the necessary documents that are needed in scope of the data management principles that have been suggested

The Data Management Plan (DMP) has been elaborated in agreement with all project partners. The DMP aims to ensure the proper handling of data collected and generated during the project's activities throughout its lifecycle. This deliverable is submitted in M6 of the project and is based on the most recent information available at the time of delivery. Nevertheless, the DMP is a constantly evolving document within the AI-MATTERS project and will be updated as necessary throughout the project's lifespan, taking into account its advancements and outcomes. The DMP will be reviewed and revised before each project management report and any updates will be reflected in these reports. Indicative stimuli of the DMP updates include the following, whereas the data management principles will always be in adherence with the FAIR principles:

- significant changes such as the availability of new data;
- changes in consortium policies;
- changes in consortium composition and external factors;
- any other reason that might be of relevance to the project

The DMP is structured in the following sections:

- Section 1 provides introductory information about the context in which the DMP, has been elaborated, its purpose, and structure.
- Section 2 presents a summary of the data to be collected or/and generated during the activities of AI-MATTERS. The purpose of data collection/generation, data types and formats, data origin, and relevant stakeholders are presented
- Section 3 describes the methodology that is applied to make data FAIR.
- Chapter 4 estimates the resources required for making the project's data FAIR, while also identifying data management responsibilities.
- Chapter 5 outlines the data security strategy applied within the context of AI-MATTERS along with the respective secure storage solutions employed.
- Chapter 6 addresses ethical aspects as well as other relevant considerations pertaining to the data collected/generated during the implementation of the project. It also elaborates on the data processing in order to assess the AI-MATTERS KPIs with particular focus on aspects related with the potential processing on personal data of the participants of activities as recommended by the External Advisor on ethics
- Chapter 7 concludes on the next steps foreseen in the framework of the project with respect to its data management plan
- Annexes include templates that will support following the indicated principles within the DMP chapters

2 DATA SUMMARY

In this chapter, a general description of data collected and used in the framework of the AI-MATTERS project is introduced together with information about data sources and owners, basic data characteristics, and use of the data sets. The main activities that are expected to collect and/or generate and process data include project management, reporting to granting authorities, service requests and execution, dissemination & communication activities, and impact assessment. Data is typically categorized based on its method of creation or capture, such as electronic text documents, spreadsheets, questionnaires, and transcripts. In addition, hereunder is an overall classification of the data types that are identified at the early stage of the project (M6) based on the relevant context for which they are generated and/or collected:

1. Project management data: such as minutes of meetings, deliverables, progress reports, monitoring of milestones, KPIs and risks logs, feedback of the Ethics Screening and Compliance Committee, Evaluation and Quality Assurance Committee, feedback from engagement with external parties such as the external Ethics Advisor, the Advisory Board members, etc.
2. Financial data: in the context of reporting the project activities to granting authorities e.g. revenue, expenses, profit margins, cash flow, and balance sheets.
3. Customer related Data: Information about customer demographics, preferences, service request & history, interactions with the TEF nodes/satellite, complaints, response times, resolution rates, and customer satisfaction surveys
4. Business activity data: including service requests forms, statistics on the services requested and executed (e.g. summary, budget, time, location, participants, etc.)

5. Technical data: data collection in real-life environments in order to find bugs or undesirable effects that were not detected by previous testing and the provision of reference data about the AI-based product's efficiency, reliability, throughput, robustness, etc., is an important point in the AI-MATTERS project. To this end, both quantitative and qualitative data will be collected including information and metadata for different scenarios of tasks that will be created to proof-test each service case indicatively production Information, production resources capabilities & capacity, sensor readings, ergonomics evaluation, feature representations, etc
6. Training programs data: such as training content, trainees profiles, etc.
7. Market and industry data: Data pertaining to market trends, industry benchmarks, competitor analysis, and economic indicators.
8. Dissemination & communication data: including material for marketing campaigns (newsletters, presentations, social media posts, press releases, banners, brochures, etc.)
9. Website and social media analytics data: user interactions, traffic data, user engagement, click-through rates, etc.

2.1 Purpose of data collection/generation

AI-MATTERS will engage in data collection in the context of the preparation of the services listed in the project catalogue, processing of the requests for service provision placed by European SMEs and larger companies, execution of the services, and the service results evaluation.

- Service requests, processing, evaluation & selection

The exchange and utilization of data will play a crucial role in the execution of service requests within the AI-MATTERS framework. These requests will be facilitated through various entry points distributed across the portals of nodes and satellites. And, additionally, the establishment of a central service desk portal to guide companies to the nodes or satellites of interest (as mentioned in T5.2), but also its integration into the AI-MATTERS website are being investigated.
To ensure efficient handling of service requests, each node/satellite will have its own committee responsible for receiving, processing, evaluating, and selecting the accepted requests. This includes launching Open Calls (Italian, Czech nodes) for specific nodes (such as the Italian and Czech nodes) and following the relevant steps, including announcement, submission, proposal evaluation, selection, funding, service execution, and evaluation.
It will be possible to submit service requests both in electronic and paper form however, following the same procedural steps. The consortium is preparing a generic template for service requests, which will be translated to the local languages of each node and satellite and be adjusted to the local regulation. Dedicated reviewers will have access to the filled in templates in order to evaluate and score them (T3.2 will detail the service request processing, service execution and delivery). Data on the requested services will be stored in terms of analyzing the requests to the nodes and estimate trends of service demand in order to prevent lost opportunities due to higher demand for services offering that is actually higher than originally anticipated. Each node/satellite leader will be responsible exclusively for the data managed in their jurisdiction and should appoint a DPO. The collected data might be used as part of key pitch element proving the impact already delivered. Data from the service execution will be collected as part of dissemination material, catalogue, and services advertising.

- Service execution & evaluation of service

Data will be collected in scope of adjusting the service to the industrial case scenario that is tested each time. Additionally, in scope of characterizing or certifying AI-solutions it will be needed to collect data on the solution's scope and characteristics, but also other data in order to assess the reliability, robustness, ethics compliance, user acceptance, as needed per case such as machine logs, sensor readings, human participant evaluations e.g. in the services that involve Human-Robot Collaboration, AR or VR tools, and so forth. It is recognized that the first customers might face difficulties without real-life operational data, so the AI-MATTERS knowledge of suitable test sites will save innovators time, shortening the time required to get the new product to market, also the provision of access to synthetic data produced by the nodes or satellite, or supporting the identification of relevant open datasets will be examined per case. All company data will be kept secure and private and all AI-MATTERS staff will operate under NDAs unless it is arranged specifically for the given case. The NDAs will be made reasonable so that non-sensitive information can appear by arrangement in AI-MATTERS publicity.

Monitoring of contracting obligations

Moreover, data will be collected to monitor the progress of the service execution and enable detecting delays as well as lack of cooperation from the applicant side.

Reporting to granting authorities

Bookkeeping activities are foreseen in scope of reporting to granting authorities thus, data such as number of days needed for service preparation for each type of service category, number of days per service execution, number of services executed, full-time equivalent per service category, total personnel costs, equipment depreciation, equipment purchases, etc. Additionally, the procedures and criteria for service requests selection will be generated in T3.2. Moreover, T2.8 will focus on establishing standard operating procedures for the service provision, where eligibility for discounts criteria will also be addressed.

Monitoring contracting obligations & service results

In terms of project management, the consortium will need to monitor and safeguard the progress of the project, which is enabled through the project KPIs including the number of companies served, number of digital and physical resources, number of cross-border access to TEF etc. as detailed in section 7.

Service optimization & alignment of service catalogue with market demand

Finally, in sake of quality assurance and service optimization customer feedback will be collected whereas a set of criteria will be defined in T2.8 for continuous improvement. Also, WP4 will perform marker analysis to support the alignment of the services with the market needs. The statistics of the requested and provided services along with customer ratings will be collected in order to focus on services that are well-received and ensure homogeneity in the service quality

Dissemination and Communication

WP5 is dedicated on raising awareness about the AI-MATTERS network and will thus generate material in scope of promoting the services and branding. Newsletters, press releases, brochures, social media posts, videos, articles on journals, and also capacity

building activities, events (physical and webinars) are among the outcomes of the dissemination and communication activities. For this scope, data from the execution of services and statistics of the service provision will be used. The datasets will be anonymized and exploited for the creation of charts, schematics, plots.

Preparation of training material for high quality service provision

T2.6 will generate material such as training content for each TEF service in order to ensure the provision of high quality services.

2.2 Data sources

The data will be collected by different means depending on the scope of their collection. Data and information about the provided services such as the available equipment, digital infrastructure and high power computing capabilities, as well as best practices and common platforms will be collected from the partners and be recorded in word, power point, or PDF formats. Subsequently, they will be published on the AI-MATTERS website.

The data collected in the service request stage or the submission of applications for Open Calls will be collected through online forms or paper forms and interviews, if necessary, submitted and stored by the particular node or satellite receiving the request. The evaluation committee of each node/satellite will be authorized to have access on these data and will be able to download the data in delimited excel and text formats. The data will be stored in excel, and word formats which are not demanding in terms of storage needs.

Quantitative and qualitative tools will be used to gather customer feedback such as online surveys, interviews, etc. (more details to be provided by T4.1). Moreover, information will be shared within the internal quality management system with a shared lesson learned tool which will be updated bimonthly. This will serve integrating the gathered feedback into the service delivery optimization. Additionally, customer rating of each service available through each node member will be collected stored and published in the AI-MATTERS website to enhance homogeneity in the quality of service provision across the network.

2.2.1 Types and formats of the collected/generated data

In accordance with Guidelines on FAIR Data Management and with ethics and legal requirements defined in the GA, all the datasets collected within the AI-MATTERS project will contain the following information:

- Service Request Identification: describing the service properties in order to facilitate the proper service request processing and selection.
- Dataset Overview: specifying what data will be collected, processed and/or generated during AI-MATTERS project.
- Dataset Format: describing the structure and type of the data, time and spatial coverage and language and naming conventions.
- FAIR: elicitation on how the FAIR principles, described in Section 3 are applied to the dataset
- Dataset Security: specifying which provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data).

Furthermore, specifying Personal Data presence and, in that case, privacy management procedures put in practice.

The provision of services will result in the generation of:

1. Monitoring and assessment of service provision results
2. Assessment of the tested solution
3. Monitoring of dissemination and communication results
4. Synthetic data might be generated and provided to the companies served, as need be

In the following paragraph, a more detailed description of the class of attributes listed above that characterize the template for the collection of datasets created within the project and attached in Annex III: Dataset Template.

2.2.1.1 Service Request Identification

First of all, it is needed to identify the service requests connected to specific datasets in terms of:

1. Company name: Name of the Company requesting a specific service and specific datasets
2. Country
3. Target AI application, where applicable
4. Target industrial sector: automotive, metalworking, and so forth .
5. Description of expected execution and outcome
6. Service category requested
7. Node that received the request
8. Status of service request (initial contact, or contract)
9. Other dataset as needed

2.2.1.2 Dataset Overview

During the service request process, various datasets may be involved, requiring each of them to be properly identified through the following attributes:

1. Dataset ID: a unique identifier for the dataset that should be a self-explaining name
2. Partner: Name of the beneficiary providing the dataset
3. Provider: Name of the provider if different from partner
4. Short description: description of the dataset in order to provide more details.
5. Data type and sources: description of the type of data, according to the list in DATA SUMMARY.

2.2.1.3 Dataset Format

A set of parameters it has been defined that can help to define the dataset format:

1. Structure description: description of the structure and type of the data. (structured | unstructured)
2. Type and format: Dataset format, specifying if it is using, for example, CSV, Excel spreadsheet, XML, JSON, etc.
3. Time coverage: if the dataset has a time dimension, indication of what period does it cover.
4. Languages: languages of metadata, attributes, code lists, descriptions.

5. Identifiability of data: reference to identifiability of data and standard identification mechanism.
6. Naming convention: description about how the dataset can be identified if updated or after a versioning task has been performed, if the dataset is not static.
7. Versioning convention: reference to how often is the data updated
8. Metadata standards: specification of standards for metadata creation (if any). If there are no standards description of what metadata will be created and how.
9. Additional comments: indication of other relevant information.

2.2.1.4 FAIR Dataset

A detailed description of how the Dataset is in line with the FAIR principles must be provided in terms of:

1. Findability: description of how the dataset can be found
2. Accessibility: description of how the dataset can be accessed
3. Interoperability: a description of standards and procedures to make dataset interoperable and to allow dataset exchange
4. Reusability: a description of procedures to be adopted in dataset re-use

2.2.1.5 Dataset Security

This section described the identified set of parameters that can help to define the dataset security aspects:

1. Personal Data (Y|N): Confirmation about personal data presence in the dataset.
2. Anonymized (Y|N|NA): confirmation if personal data is anonymized.
3. Secure storage procedures: Information about how managed data recovery and secure storage was.
4. Encryption (Y|N): identification of encryption procedures
5. Dataset ethics and legal requirements: describe any Ethics and/or legal requirements relevant to the dataset if not covered in previous questions.

2.2.2 Collected dataset

The goal to collect all the information, described in the previous paragraphs, has been achieved through the creation of a shared template (Annex III: Dataset Template), which will be the reference to collect all datasets that will be used in the project . It has to be edited by the Consortium directly on the Project repository (or at least the changes should be uploaded by each partner on the master document) and every time a new dataset is produced and used.

In the next version of the Data Management Plan, the dataset collected and generated during the project will be listed.

2.3 Relation of project objectives and data collection

The data collected will help prepare, execute, evaluate, and improve the services of the TEF. This fits the following project objectives:

I.3- Scope and service portfolios

I.4- Service provision management
II.1- Quality Assurance
II.2- Readiness and Completeness
II.3- Outreach, awareness making and Branding
II.4- EU-wide Synergy and Complementarity

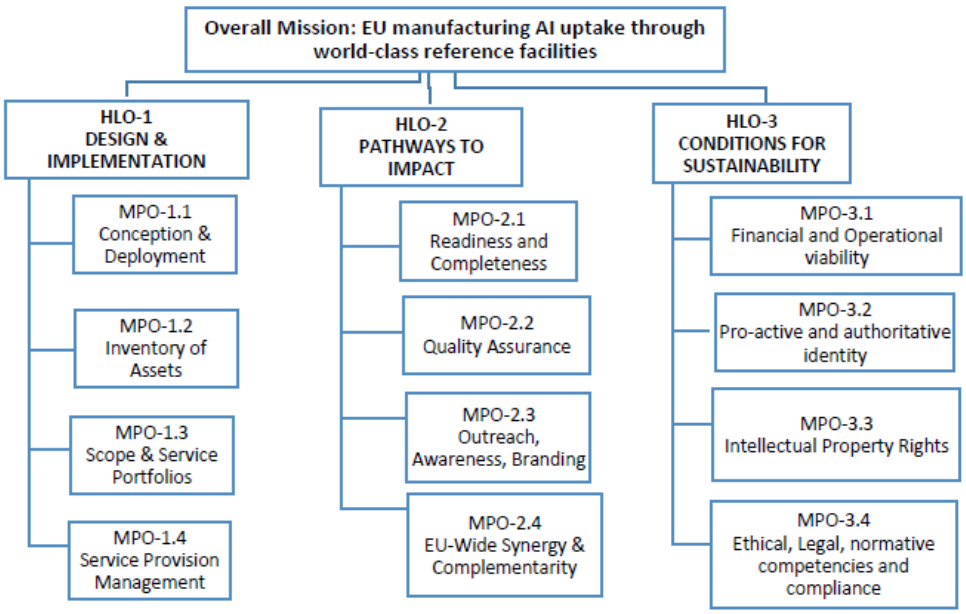

Figure 1: AI-MATTERS Hierarchical Project Objectives (overarching, High-level and Project Objectives)

2.4 Data security

Personal data collection will be limited in scope to the service request submission and informed consent of participants about the use of personal data will be required. All the collected data will be stored in servers located in EU. Some of the data to be tracked are considered personally identifiable according to the European GDPR framework (such as location, Internet Protocol address, name and surname) In order to comply with regulation appropriate steps will be taken towards anonymising such data (e.g. UserID across devices anonymization, IP anonymization, tracking data logs deletion etc.). Users will be provided with the right to deactivate their accounts in case they wish to delete their data. This will result in deleting their personal userID and thus, all personally identifiable data assigned to the userID in question. If users wish the AI-MATTERS consortium to delete, suspend, or correct their data, or if they wish to receive information on the saved data relevant to them they may contact the responsible AI-MATTERS consortium party (i.e. the corresponding node leader) with a relevant request. The AI-MATTERS beneficiaries will use economically justifiable efforts to honour the requests and might retain an archived copy of the users; records if required by mandatory law.

All datasets will be anonymised, prior to their use, for research and impact assessment. All company data will be kept secure and private and all AI-MATTERS staff will operate under NDAs unless it is arranged specifically for the given case. The NDAs will be made reasonable so that non-sensitive information can appear by arrangement in AI-MATTERS publicity. Moreover, consent forms will be distributed prior of sharing data related to a particular company and the associated service execution and results.

The Advisory Board that will provide advice on challenges and topics related to the project scope will operate under the NDA that has been a part of the Annexes of the CA, whereas

the Executive Board will ensure that the NDA is executed between all parties and each Advisory Board member.

If the situation arises that data needs to be transferred to entities outside the EU, AI-MATTERS will seek the approval of the competent Data Protection Office before any transfer occurs (see also Section 2.3.1). The only exception will be in the case of countries which are on the list of countries which provide adequate safeguards with respect to the protection of privacy and fundamental rights and freedoms of individuals. All personal data protection processes will adhere to Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data as well as Directive 95/46/EC (GDPR - General Data Protection Regulation).

2.4.1 Reuse of existing data

An important point in the overarching objective of AI-MATTERS is to allow for testing of AI solutions and optimization under real industrial scenarios in order to refine performance, efficiency, reliability, throughput, robustness, etc. Providing access to relevant data is a priority for AI-MATTERS consortium that will investigate among others reuse of open datasets, synthetic data sets produced by the consortium members, or establishing procedures in order to exploit data that have been produced during the execution of experiments. In particular, T2.8 aims among others to establish common rules for dealing with IP, T3.3 will cover the innovation and IPR management strategy along with a methodology for managing all background and foreground knowledge, general terms of use and relevant IPR provisions, joint exploitation plans for the consortium and the means and procedure for exploitation of assets .e.g data. Standardized NDA is currently investigated for the management of IPR, whereas the ownership of the created know-how and the reuse of the gained know-how for other clients will be meticulously reviewed within T3.2 that is also working on IPR management issues.

Data will be collected throughout the project, but no datasets of previously collected personal data will be used or merged for the purposes of the project. Likewise, sharing data with non-EU members not being part of the consortium is not foreseen.

2.5 Data utility

The data will be exploited by AI-MATTERS project partners and companies served by the TEF for six primary purposes:

1. Evaluation of service requests
2. Preparation & execution of the service
3. Research
4. Results assessment
5. Impact assessment
6. Services continuous improvement

Service evaluations are to be carried out by dedicated committees at the nodes/satellite level. Each committee will be granted limited access to a specific number of data registers

from the evaluation data set. All members of the committee (and also evaluators of Open Calls at the applicable nodes) will be required to sign and adhere to the following:

1. Code of conduct for evaluators

2. Confidentiality and no conflict-of-interest declaration

Only when the above documents are signed will data access be granted to the evaluators. This will be for a limited time, online, through secure authentication provided by the portal of the relevant node or satellite.

Anonymised data sets will also be made available to third parties on a case-by-case basis in scope of dissemination and communication, or project evaluation. Additionally, anonymized service data collected by service execution might be of use for future service provisions. The possibility of providing such data is investigated by the consortium, whereas the terms of proceeding with such action are to be specified in T2.8, T3.2, and T3.3.

2.6 Intellectual property rights

AI-MATTERS will follow the general principles for intellectual property rights (IPR) as described in the "Model Grant Agreement for the Horizon 2020 Framework Program". Project partners and third-party beneficiaries are individually responsible for taking the appropriate measures to protect their IPR created during the project. T2.8, T3.2, and T3.3. will be dedicated on handling IPR topics.

All intellectual property resulting through work carried out by third party beneficiaries will remain their property. The ownership of data generated through the service execution will be examined per case depending on the sensitivity of the data for the company that is served, the ownership of data that were used as input for the experiment, the ownership of the equipment or algorithm generating the data, etc.

Any intellectual property produced by any of the partners during the project will be licensed, free-of-charge, to other project partners for the duration of the project, if needed for the proper execution of the project. Third party beneficiaries are under no obligation to provide access to project results (but are, of course, free to do so if they so choose). In all remaining cases, all relations between members of the Consortium will be defined by the Consortium Agreement and the applicable (if any) national laws.

2.7 Research data and Open Access

One of the goal of the AI-MATTERS project is fostering access to and re-use of generated data ensuring to make research data as open as possible and as confidential as necessary.

The Consortium will identify an open access repository both for scientific publications and research data. One of the candidates is the ZENODO Platform, hosted by CERN which has existed since 1954 and currently has an experimental programme defined for the next 20+ years.

The project coordinator or the dissemination leader will create an account in order to upload respective Open Access publications, Open Access datasets or a validated public deliverable.
Publications and data uploaded on Zenodo are indexed automatically in the OpenAire aggregator and will be automatically synchronized with the European Commission participant portal, hence there is no need to update information on scientific publications for reporting purposes.
In addition to Zenodo, project partners might want to use other platforms to upload data or publications: further repositories for data are available within the "re3data" registry (www.re3data.org) and OpenAire (www.openaire.eu). Project partners are encouraged to disseminate outputs also in web-based scientific social networks such as www.academia.org or www.researchgate.net.

3 FAIR DATA

AI-MATTERS will manage the data from the provided services, as well as the relevant promotional, impact assessment, and service assessment results. The collection of data through an online application form, when possible, will facilitate data management by having the information of third parties structured in a standardised form. The FAIR[1] guiding principles for scientific data management and stewardship will be used throughout the project. As per the FAIR principles data should be:

1. Findable
2. Accessible
3. Interoperable
4. Reusable

3.1 Data discoverability

The data management plan supports the effective collection and integration of AI-MATTERS project data. Storage, processing, and sharing will occur via the CEA proposal submission platform, and different events and meetings. To be able to evaluate the success of the project and its contribution to the AI and robotics adoption, AI-MATTERS needs to be capable of establishing a relation between provided services, topics covered, and the achieved results. Data will be anonymised, that is, data will not identify individual persons.

To be able to use the data generated by the project it is essential to integrate data from the participants and the activities undertaken by project partners. Considering the FAIR data principles (meta)data should:

[1] Wilkinson, M. D. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

- Be assigned to a globally unique and persistent identifier
- Contain enough metadata to fully interpret the data, and
- Be indexed in a searchable source

By applying these principles data becomes discoverable and includes authentication and authorisation details.

3.1.1 Data identification mechanisms

All documents associated with a service will be identified with a unique, and persistent, numeric identifier that will be given at the time of the submission process.

Examples:

- 001RequestForm
- 001TopicsDescriptionIdentification
- 001RequiredEquipment
- 001Ethics
- 001EvaluationReport

As per the documents related to project activities and/or deliverables, the tasks or deliverables number will be used to identify the document followed by a brief title of the activity or deliverable.

3.1.2 Naming conventions

The guidelines for the naming conventions are summarized in the list below:

- File names should be kept short but meaningful
- Unnecessary repetitions and redundancy in file names and file paths should be avoided
- Capital letters should be used to delimit words
- In the case when numbers are to be used in a file name, two-digit numbers, i.e. 01-99, should be used except for the case of years or other numbers with more than two digits
- When a date is used in the file name it should be stated 'back to front', and use four-digit years, two-digit months, and two-digit days: YYYYMMDD, or YYYYMMDD-DD, or YYYYMM, or YYYY, or YYYY-YYYY
- When including a personal name in the file name the family name should be given first followed by the initials
- Words such as "draft" or "letter" should be avoided at the beginning of the file names, unless doing so will make it easier to retrieve the record
- Care should be given into ordering the elements of a file name in the most appropriate way to retrieve the record
- The file names of records relating to recurring events, e.g. minutes of periodical meetings, should include the date and the description of the event, except where the inclusion of any of either of these elements would lead to unnecessary repetitions and redundancy

- The file names of correspondence should include the name of the correspondent, an indication of the subject, the date of the correspondence, and the statement of whether it is incoming or outgoing correspondence except where the inclusion of any of these elements would lead to unnecessary repetitions and redundancy
- The file name of an e-mail detachment should include the name of the correspondent, an indication of the subject, the date of the correspondence, “attch”, and an indication of the number of attachments sent with the covering e-mail, except where the inclusion of any of these elements would lead to unnecessary repetitions and redundancy
- The version number of a record should be indicated in its file name by the inclusion of “v” followed by version number and where applicable “draft”
- Non-alphanumeric characters should be avoided in file names

Regarding the naming of deliverables, they following nomenclature should be applied:

Working version: Dx.y_YYYYMMDD_vz.w
Final version: Dx.y_YYYYMMDD_vz.0

Adopting a consistent set of naming conventions when developing the project's data files can significantly enhance their discoverability and facilitate working with multiple files. With this in mind, AI-MATTERS employs consistent data file names that include references to their content, status, and versioning, while also increasing their searchability. By doing so, both project partners and interested stakeholders can readily identify a file, classify and organize them as needed.

According to the UK Data Archive creating brief yet meaningful names for data files that facilitate classification is a best practice in naming conventions. The naming convention should avoid the utilisation of spaces, dots and special characters (such as & or !), but rather use underscores to separate elements in the data file name and make them more comprehensible. Additionally, versioning should be part of the naming convention to clearly identify changes and edits made to a file.

To facilitate the reference of the datasets produced during the project's implementation, AI-MATTERS will employ a standard naming convention that incorporates versioning and supports the possibility of creating multiple datasets during an activity that involves data collection or generation. AI-MATTERS's naming convention takes this issue into account by employing a unique element that captures the number of datasets generated during the same activity.

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

AI-MATTERS _ [Name of Service] _ [Number of dataset] _ [Issue Date] _ [Version number]

- AI-MATTERS: The name of the project.
- Name of Service: A short version of the name of the service 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 (YYYY.MM.DD.).
- 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 AI-MATTERS are provided below. These examples aim only to showcase the naming conventions and do not necessarily correspond to actual datasets.

- AI-MATTERS_NeedsAndRequirements_Dataset1_2023.10.30_v1 – The first dataset generated within the framework of the survey conducted to identify the needs and requirements of the diverse AI and Robotics stakeholders. This is the first version of the dataset that was last modified on the 31st of November 2023 (30/10/2023).
- AI-MATTERS_HRCAssessment_Dataset2_2023.12.15_v2 – The second dataset created in the process of assessing a Human Robot Collaborative solution. The last modification of this dataset, which in this case produced the second version of the dataset, was on the 15th of December 2023 (15/12/2023).

3.1.3 Keywords

Documents that are related to the project activities will be based on templates agreed upon by the consortium, which will include a keywords section to make documents discoverable. Also, the service requests and the applications submitted in open calls (for the nodes where open calls are applicable) will use keywords related to the topics covered by the services of AI-MATTERS. The keywords used to easily identify documents related to a specific service request will be the ones used throughout the request processing (where applicants will have to select the characteristics of their service using the descriptors from a drop down menu. It will be possible to export an excel spreadsheet with all the information about the service requests in sake of reporting and periodically evaluating the alignment with market demand, and ensuring homogeneity of the services quality at the project level.

Furthermore, the project's data will be provided with search keywords with a view to optimizing its re-use by interested stakeholders during its entire lifetime. With that in mind, the metadata standards employed by AI-MATTERS 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 AI-MATTERS, 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 will comprise terms related to human robot collaboration, robots, machinery safety systems, software systems as well as augmented reality. The keywords will accurately reflect the content of the datasets and avoid words used only once or twice within them.

3.1.4 Document versioning

Only documents created with the consortium will be versioned. as a result document templates including a history panel, such as the one is illustrated in Table 2, and three status descriptors (each one associated with each step of the life cycle of a document) will be used:

1. Draft: This document version is under development by one or more partners
2. Under review: This document version has been sent for review by one or more partners
3. Published: This document version has been submitted to the Commission Services

Table 2. Document template extract showing the history panel

Version	Issue Date	Stage	Comments	Contributor
0.1	DD.MM.YYYY	Draft/Under review/Published	Lorem Ipsum	XX, YY
0.2				
0.3				
1.0				

Project partners, following the recommendations included in Section 3.1.2., will identify different document versions by using a two-digit number following the descriptor “Draft”. A document reviewed by another partner should be returned to the principal author including “Edit+Initials” of the secondary author(s)/editor(s), where the “Edit” will be marked with track changes, where applicable, or highlights and comments. Only the principal author will change the draft number and will add the term Final to documents ready to be sent to the EU (or those to be used as final versions by the Consortium internally). The document history included in the document template should be filled in as shown in Table 3.

Table 3. Document template extract showing the history panel

Version	Issue Date	Stage	Comments	Contributor
0.1	DD.MM.2023	Draft	Section 3.1. needs to be completed	Jane Smith
0.1	DD.MM.2023	Under review	Section 3.1. completed. Comments added in the document	John Smith, Jane Smith
0.2	DD.MM.2023	Draft	Added suggestions by John Smith	Jane Smith
1.0	DD.MM.2023	Final	Final version with the contribution of the involved partners	Jane Smith

3.1.5 Metadata standards

Basic metadata will be used to facilitate the efficient recall and retrieval of information by project partners and external evaluators. Additionally, metadata will facilitate the discoverability of the requested information. In this view, all documents related to the project have to include in the front-page an information panel with information about the author(s) and contributor(s), WP, dissemination level, and nature of the document (see Figure 3).

Project Acronym: AI-MATTERS
Project Title : AI MAnufacturing Testing and experimenTation network For EuROpean industrieS
Date of Delivery – Contractual: MX
Date of Delivery – Actual: MY
Version: X.X
Nature: YYY (see types in the list hereunder)
Dissemination Level: YYY (see types in the list hereunder)
Lead Beneficiary: XXXX
Reviewer(s): Affiliation
Keywords data management, data protection, data privacy, data security

The "Dissemination level" in the document information panel (see Figure 3) has three possibilities:

- PU –Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)
- SEN –Sensitive, limited under the conditions of the Grant Agreement
- Classified R-UE/EU-R –EU RESTRICTED under the Commission Decision No2015/444
- Classified C-UE/EU-C –EU CONFIDENTIAL under the Commission Decision No2015/444
- Classified S-UE/EU-S –EU SECRET under the Commission Decision No2015/444

The "Type" in the document information panel (see Figure 3) has four options:

- R: Document, report (excluding the periodic and final reports)
- DEM: Demonstrator, pilot, prototype, plan designs
- DEC: Websites, patents filing, press & media actions, videos, etc.
- DATA: Data sets, microdata, etc.
- DMP: Data management plan
- ETHICS: Deliverables related to ethics issues.
- SECURITY: Deliverables related to security issues
- OTHER: Software, technical diagram, algorithms, models, etc.

Additional metadata will be generated by the service requests (and Open Calls application, at the applicable nodes) process. The data included in the service requests/application forms by Open Call project applicants will be used to identify documents and make data discoverable. The service requests will be submitted either

online at the node platform that is selected by the applicant, or in person via filling in written applications. Each node or satellite is exclusively responsible for managing the data associated with the service requests submitted to it and service execution held at its facilities.

3.2 Data accessibility

3.2.1 Openly available data

The full set of anonymous data that have been shared in the context of dissemination and communication activities will be made available to third parties that request access to it for research purposes. The requests for access to data sets generated from testing and experimentation activities will be processed per case, whereas the nodes and satellite should strive to make research data as open as possible and as confidential as necessary. Furthermore, the anonymised datasets will be exploited through the creation of tables, and infographics that will be updated periodically and/or at the end of the Open Calls (Italian and Czech nodes). These generated tables and infographics may be used as part of the dissemination activities of the project.

Data will be collected throughout the project, but no datasets of previously collected personal data will be used or merged for the purposes of the project. Likewise, sharing data with non-EU members not being part of the consortium is not foreseen.

3.2.2 Procedures to access data

The data sets will be made available through the AI-MATTERS website, as well as the platform of each node/satellite. No specific software tools will be needed to access the data. The anonymised data sets will be saved and stored in word, pdf or excel formats to facilitate their exploitation (as well as guarantee their long-term accessibility).

3.2.3 Handling access to restricted data

The platforms of the nodes and satellite, as well as the AI-MATTERS website should be equipped with the necessary tools to handle data associated with them at multiple levels of restrictions. In particular the following capabilities should be made possible by the tools in question:

1. The creation of users with different access levels
2. Access for specific time frames
3. Access to specific data sets
4. Access through secure authentication

3.2.4 Data repository

The nodes and satellite will collect data obtained through the service requests and Open Calls (at the nodes where this procedure is applicable), on their individual IT infrastructures. Additional data that will be generated in scope of project management by project partners and third parties will be deposited on the internal project cloud infrastructure “Talkspirit” (provided by CEA).

3.3 Data interoperability

3.3.1 Ascertainment of data interoperability

The project partners will be responsible for storing all data in the appropriate format that will make data accessible to all professionals that might be interested in exploiting the data generated during the project. The capacity of AI-MATTERS generated data to be interoperable by third parties will be assessed through several reviews during the project. The outcome of these reviews will be the guarantee that AI-MATTERS generated data will be compatible with the needs of specific end users.

Best practices and standards for cloud interoperability have been proposed recently, such as the Open Virtualization Framework, the Cloud Infrastructure Management Inter- face, SWIPO (Switching Cloud Providers and Porting Data), or the standard ISO/IEC 19,941:2017. Promising initiatives like the European project GAIA-X, which aims to establish a federated data infrastructure by integrating cloud and edge services and required data centers across Europe, might boost AIaaS diffusion and provide the foundation for seamless integration of AIaaS and exchange of data between providers and user[1] to prevent invasion of individuals' privacy when collecting and gen erating data about them and to allow users to understand the consequences of data disclosure better. membership inference attack. The AI-MATTERS acknowledges such initiatives and will be evaluating their integration into the project activities.

Extant research has started to propose several approaches to protect people's privacy during the training and operation of an AI in the cloud, such as training AI models using encrypted data, making encrypted predictions, as well as returning the predictions in an encrypted form. third-party attestations and related certifications are promising means to assess whether an AIaaS has implemented adequate data governance and protection mechanisms.

3.3.2 Vocabulary used

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 data. Therefore there must be a shared understanding of the data's contents, context, and meaning, in order to facilitate the smooth transfer of data between different systems and services. In this scope, AI-MATTERS adopts the Dublin Core Metadata standard that includes a set of 15 specific metadata elements that address quality and consistency requirements while still allowing for interoperability with other data sources in the linked data environment (Table 4).

Table 4. Dublin Core Metadata Standard Vocabulary

No	Element	Element Definition
1	Title	A name given to the resource.
2	Creator	An entity 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.

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	Refers to the language in which the data is written or presented
13	Relation	Identifies a relationship between the data and other data resources.
14	Coverage	Specifies the spatial or temporal aspects of the data.
15	Rights	Specifies any legal rights associated with the data, such as copyright or patent information

Additionally, the interoperability of open data will be facilitated by Zenodo, as its metadata is stored in a JSON format following a defined JSON schema that includes HTML microdata, which allows machine-readable data to be embedded within HTML documents as a series of nested name-value pairs. The JSON schema offers a set of shared microdata vocabularies that can be used to markup pages in a way that is easily understood by major search engines.

3.4 Enhancement of data reuse

Clauses referring to Access Rights and Non-Disclosure of Information (see Section 9 and section 10 of the Consortium Agreement) are included in the Consortium Agreement (CA). These clauses, as well as Section 4.4 of the CA (which offers a description of the approach of the project in ensuring conformity to GDPR), will be key features governing the use of data by third parties.

Information related to the funded SMEs, and larger companies (such as the company name), will be published, for dissemination purposes only, after obtaining the consent of the beneficiary, or without such consent if publication is required by EU rules. Regarding data produced by sponsored projects, each beneficiary will be responsible for permitting or restricting access to their data and results

3.4.1 Data use by third parties

Three data access pathways for third parties are envisaged. Firstly, external evaluators will be granted access to a restricted number of registers from the data sets during the evaluation phase of each Open Call (applicable to the Italian and Czech nodes). To be granted access to data evaluators will be required to sign an "Expert Evaluator Code of Conduct", an "External Evaluation Fundamentals", and a "Declaration of Confidentiality and No Conflict of Interest" documents. Access will be online, for a limited period, using secure authentication secured by the nodes that organize the Open Calls.

Secondly, during the project, third parties will be allowed access to anonymised data sets on a case-by-case basis (see Section 3.3.1). The provision of project data that have been generated by the execution of services to other parties will depend on the data owner If a member of the consortium is the data owner, they should follow the principles of making research data as open as possible. If the owner is a company that has received a service the data use by other third parties will depend on the use that third parties want to make of it, and the added value of sharing such data, as well as the terms of the NDA signed between the company and the corresponding node or satellite. Lastly, in the same line as described, after the end of the project, the openly available generated data will remain available (for access by third parties) for four years (see Section 3.4.2) on the individual platforms of the nodes/satellite and/or the AI-MATTERS website.

3.4.2 Data availability for reuse

The Consortium will maintain project data in a reusable state for as long as possible after the end of the project; an initial availability period of four years is foreseen (this time may change depending on agreements among the project partners). Data related to the open calls will be made accessible after each round (once the final list of beneficiaries is published), whereas data related to the service delivery will be made periodically available – the period is a topic of discussion to be agreed upon in the next consortium meeting. All other data generated by the project will be released in agreement with the participating entities once the project comes to its end.

3.4.3 Data quality assurance

The project coordinator, assisted by the Quality Assurance Committee, the Executive Board, as well as LMS (leader of T1.3), IPA (leader of T2.8) and experts of the Advisory Board will be responsible for assuring the quality of the data by making sure that datasets follow the FAIR principles included in this plan, and that data is updated on a regular basis.

Personal data processing will be done in accordance with EU, national, and international laws considering the "data quality" principles listed below[2]:

- Data processing is adequate, relevant and non-excessive
- Data are accurate and kept up to date
- Data are processed fairly and lawfully
- Data are processed in line with the rights of data subjects
- Data are processed in a secure manner
- Data are kept for no longer that necessary and for the sole purpose of the project

The data quality assurance process will be in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data.

[2] Wilms, G. Guide on Good Data Protection Practice in Research of the European University Institute. (March 2017). Retrieved from http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf on 4 June 2019.

4 RESOURCE ALLOCATION

4.1 Data management responsibility and related costs

The nodes and satellite will be responsible for managing exclusively the data of service applicants stored on the individual platforms of each. Project management data will be stored on Talkspirit, and CEA will be responsible for the overall management.

Regarding the data resulting from the activities of the project, each WP leader will be responsible for the storage, and compliance of the data, and its uploading on the AI-MATTERS Talkspirit platform, or other storage systems.

LMS assisted by the WP leaders and node leaders will be responsible for updating the DMP and developing a strategy to encourage:

- The identification of the most-suitable data-sharing and preservation methods
- The efficient use of data. assuring clear rules on its accessibility
- The quality of the data stored
- The storage in a secure environment with a user-friendly interface.

No extra costs, apart from those linked to the maintenance of the AI-MATTERS website, Talkspirit and nodes/satellite platforms are expected for managing data and making sure it abides by FAIR principles. Each consortium partner is solely responsible for the costs of processing data they possess.

4.2 Costs estimations for long-term data preservation

No additional data storage and maintenance funding will be required once the project ends. It is important to consider that the topics covered by the project correspond to current AI and robotics testing and experimentation needs. The impact, relevance, and value of project results will be dependent on the evolution of challenges faced by end users.

5 SECURITY

The node leaders and satellite leader will collect data from companies requesting services through their AI-MATTERS portals or in paper form. Each of the node leaders and the satellite leader will be exclusively liable for the data under their jurisdiction. This data will be used during the service request processing by the evaluation and selection committees of the nodes/satellite (more details to be provided within T3.2) or the Open Calls evaluators (Italian and Czech nodes). Data will be deposited, and secured, in the platforms of the node and satellite leaders. The anonymised data will be accessible through the web interfaces of the platforms in question in delimited text, excel, word, and PDF formats. Access to the data sets will only be available to authorised users via authentication.
The platforms of the node leaders and the satellite leaders should apply technological and organizational measures to secure the processing of personal data against access by unauthorised persons, processing in violation of the law and change, loss, damage or destruction. These measures cover issues such as:

1. Information security
2. Data access options
3. Backups
4. Accidental deletion or modification of data
5. Deletion or modification of data by users
6. Any other issues that will be identified by the consortium as the project activities advance

The node leaders and satellite leaders should have specific terms and conditions for use of their platforms that must be accepted by users of the platform prior to any form of access being granted.

6 ETHICS

6.1 Data related ethics screening aspects

AI-MATTERS has appointed an external Ethics Advisor who has already analyzed the ethics issues of the project and has provided a relevant report. The EA has signed a "Declaration of Confidentiality and No Conflict of Interest" and has proven experience and competence in analysing ethical issues relating to privacy and data protection (particularly in the fields of AI and robotics). In parallel, AI-MATTERS will appoint an Ethics Screening and Compliance Committee that will be made up of ethics experts of the consortium parties: The ESCC will perform the following activities under the guidance of the EA:

- Raising awareness, throughout the project, of potential ethical issues regarding privacy/confidentiality of personal data
- Evaluating the accomplishment of ethical requirements of third-party beneficiaries, particularly during the selection process
- Monitoring ethical concerns related to sponsored projects
- Offer potential solutions and contribute to the resolution of conflicts that may arise

The Coordinator, with the support of the EA, the ESCC, and the General Assembly will ensure the correct application of H2020 standards, and guidelines, by all participants regardless of the country in which activities will be carried out. This will be achieved through:

- Ensuring that the provisions on ethics in the H2020 Regulation and in the Rules for Participation are respected (article 19 of H2020 Regulation)
- Ensuring that the European Code of Conduct for Research Integrity (2017) is adhered to
- Taking preventive and/or corrective measures when necessary
- Addressing issues related to breaches of research integrity and misuse of results
- Encouraging the design of project activities following ethical principles

Ethics issues are discussed in detail in Deliverable D1.7: Ethics Management plan, which will be updated annually (D1.9 [M12], D1.10 [M24], D1.11 [M36], D1.12 [M48], D1.13 [M60]), whereas T1.5 is dedicated to ethics screening and compliance. Additionally, T3.1 contributes in ethics screening, T2.7 will establish rules to ensure the ethical use of AI, T4.3

will work on ethical issues as well, whereas T1.4 will ensure transversal liaison and synergies towards competent authorities for ethics screening.

The key areas to be monitored by the aforementioned stakeholders and the partners involved in the mentioned tasks have to do with interactions in environments where humans and robots co-exist.

6.2 Personal data processing

In addition to the above, the AI-MATTERS project will implement Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

The project will also ensure that the procedure for informed consent with regards to the collection, storage, and protection of personal data, is implemented whenever this is necessary. The project will implement the highest ethics standards, and criteria, in identifying/recruiting research participants, and the informed consent procedures that will be implemented for the participants. Templates for informed consent forms and information sheets will be created and used whenever necessary.

The AI-MATTERS project will follow all relevant national, and EU, safety standards regarding safety issues related to activities involving human-robot collaboration experiments. The consortium will ensure that all appropriate health and safety regulations and procedures (as per local/national/EU standards/guidelines/legislation) are adhered to by all individuals involved in the project. All documentation required by local laws and regulations regarding ethics approval of human-robot collaboration experiments will be provided to all participants.

Personal data will be protected in the following ways:

1. Unnecessary personal data collection will be avoided

2. The personal data needed for statistical analysis will be collected anonymously

3. Personal data will be collected only within the legal basis outlined in Article 6 of the GDPR

4. Personal data will be treated in strict confidence

5. Personal data will be protected using appropriate technical means

6. Individuals will have the right to access their personal data,

7. Individuals will have the right to withdraw their personal data

Horizon 2020 rules on ethics issues will be adhered to and, in case of conflict with the national/local ethics rules and regulations, they will prevail. Lastly, the project will closely follow the latest developments in the field of Roboethics[3]. Compliance to the respective national legal frameworks should be ensured by the consortium parties. Also, AI-MATTERS will monitor the procedures for the adoption of AI Act regulation and adjust the project procedures accordingly, if necessary.

[3] http://www.roboethics.org/

6.2.1 Anonymisation/pseudonymisation techniques

Anonymization and pseudonymization techniques should be implemented in order to enhance measures to protect protect sensitive or personally identifiable information (PII) by transforming or obscuring the data in a way that individuals cannot be directly identified.

Anonymization techniques: Anonymization refers to the process of removing or altering PII in a dataset so that the data can no longer be linked to an individual. The goal is to prevent the data from being reasonably re-identified. Indicative techniques include:

- Data aggregation: Combining data to form groups or categories, making it difficult to identify individuals within the aggregated data.
- Generalization: Removing specific details or replacing them with broader, less identifying information. For example, replacing exact birthdates with age ranges.
- Randomization: Introducing randomness into the data, such as adding noise or replacing values with random equivalents, to make it challenging to trace back to individuals.
- Data masking: Masking or redacting specific data elements that could identify individuals, such as names, addresses, or social security numbers.

Pseudonymization techniques: Pseudonymization involves replacing or encrypting identifiable information in a dataset with pseudonyms or unique identifiers. The pseudonyms are distinct and unrelated to the individuals' actual identities, but can be used consistently to track and link data within the dataset.

- Tokenization: Replacing sensitive data with randomly generated tokens, preserving the structure and relationships within the data.
- Encryption: Transforming the data using cryptographic algorithms, requiring authorized decryption to retrieve the original information.
- Hashing: Converting data into fixed-length strings of characters using hash functions, making it computationally infeasible to reverse the process and obtain the original data.

The choice of technique should be made based on the specific privacy requirements, the nature of the data, and the intended use of the anonymized or pseudonymized dataset.

6.2.2 KPIs & Personal data protection

In the context of the KPIs stated in the DoA, two have been pointed out by the EA to need further elaboration on aspects of personal data protection. In particular:

- POL-KPI-2-1: Number of aged individuals received training services from the TEF (> 220),
- POL-KPI-2-2: Number of cases in which aged workforce is assisted by AI and/or robotics solutions tested in TEF (> 150).

It is acknowledged by the project consortium that the information provided in the DoA on the aspect of addressing the problem of aging workforce and the wording used in the KPIs is misfortunate and does not correspond well with the actual intensions behind the approach undertaken in the project, which has been aimed at ensuring inclusiveness and equalisation of opportunities for all members of the manufacturing industry community,

regardless of their age, gender or any other social or personal characteristics. It has also been declared that the AI-MATTERS consortium will take their best efforts to provide measures in the project to prevent the risk of any type of discrimination such as the following when selecting participants for the service execution e.g. for testing Human-Robot Collaboration solutions, AR or VR based solutions, and so forth:

- Inclusive Criteria will be developed in T2.7 for the participant selection based on relevant factors for the research objectives, rather than characteristics that may lead to discrimination. The criteria will be fair, transparent, and will not exclude individuals based on protected characteristics such as race, gender, age, disability, or religion.
- Stereotyping will be avoided when selecting participants. Each potential participant will be treated as an individual and their suitability will be assessed based on relevant qualifications, experience, or characteristics directly related to the research objectives.
- Ethical Approval and Review will be ensured by the ESCC and the EA.
- Informed Consent Process will be implemented (see Annex I and Annex II) including the right to withdraw from the study at any time, without facing any adverse consequences.
- Anonymous or Pseudonymous Data Collection will be performed whenever possible and appropriate for the research objectives to protect participants' identities and reduces the risk of potential discrimination based on identifiable information.
- Diversity and Representation among research participants will be strived to ensure a broad range of perspectives and experiences. Participants from different backgrounds, cultures, genders, ages, and socioeconomic statuses, among other relevant characteristics should be included in the testing.
- Regular Monitoring and Evaluation will be performed by the relevant governing bodies, the EA, and in the context of the relevant project tasks (T2.7, T1.5, T4.3 and T1.4).
- Training and Sensitization will be provided to the research team members in the context of T2.6 “Staff assignment and training for the service provision”.

7 Conclusions

The current version of the Data Management Plan (DMP) outlines the proposed methodology to be implemented by the AI-MATTERS project in order to ensure the efficient and effective management of the data generated and collected during the project's activities. The methodology outlined in the DMP has been crafted to ensure the FAIR principles are upheld throughout the lifecycle of the data and to address the security and ethical concerns that accompany the collection and generation of the data.

It is anticipated that the DMP may be updated as the project evolves to reflect the latest developments and outcomes. This living document will be reviewed and revised prior to each periodic project management report, with any modifications being incorporated into the reports as necessary.

Annexes

Annex I: Information sheet

Information Sheet

Logo of organization responsible for the experiments and with appointed DPO	<Full name of organization (node leader/satellite leader)> <Address> <Tel: > <http://www.xxx.xxx>

AI-MATTERS OBSERVATION

Purpose of Research

You have been invited to participate in this study that is part of the AI-MATTERS project because [explain here how researcher came into contact with the respondent, for example because respondent is registered with a panel]. The study isperformed under the service entitled XXXXX <name of the service> in order to investigate YYYYYYY <scope of service.

Before you decide to participate, we would encourage you to read the following information, so that you know what the study is about, what is expected from you and how we deal with processing your personal data. Given this information you can indicate using the consent form, which is provided together with this information sheet, whether you consent to take part in this research project and the processing of your personal data.

Do I have to take part?

Your participation is entirely voluntary. We would like you to consent to participate in this study, as we believe that you can make an important contribution to the research. You are free to withdraw from the project at any point without any explanation. You can also withdraw your data for a period of up to 10 days after the interview, as after this time it will not be possible to identify your individual data from the aggregated anonymised results.

If you are uncertain or uncomfortable about any aspect of your participation, please contact the researcher at the end of this information sheet to discuss your concerns or request clarification on any aspect of the study.

What will I do if I take part?

The research involves observations by one or more investigators. It is possible that the investigators observe you while <YYYYYY description of activity>. These activities will be

recorded using photography or videorecording. The session should take about <xx minutes>. All observations will be recorded, and we may use the results in publications and presentations related to this project. <If any additional equipment is being used, please describe it: e.g. physiological, eye tracking equipment> If you feel uncomfortable during any moment, please let the researcher know, and we will adjust or remove the equipment.

What are the possible benefits of taking part?

We hope that with your help and that of other participants, we will get an understanding of how AI and robotics technologies can contribute to future improvements of manufacturing working conditions, safety, and profitability. Also, we hope that you will find the experience fun and interesting. <(If applicable) To thank you for taking part you will receive xxx>

How will my data be used?

Any information you supply will be treated confidentially in accordance with the General Data Protection Regulation (EU GDPR) and the <National legal framework> to protect your privacy. Anonymised transcripts will be created by removing or replacing identifiers such as name, age, and location. Under no circumstances will identifiable responses be provided to any third party. Your interview recording and transcript will be securely stored on a password-protected hard drive at <Node/satellite leader facilities> and will be deleted as soon as they are no longer required. We will not use your name when we write up the results, but we would like to use some quotes of what you say. Findings of the study might be published in academic journal articles, presented at academic conferences, at university seminars, or on social media.

Questions About the Research?

If you have questions about the research in general or about your role in the study, please feel free to contact<researcher conducting the study and their email > .This study has been approved by the AI-MATTERS Ethics Screening and Compliance Committee. For more information on the AI-MATTERS project please visit https://ai-matters.eu/

Thank you for reading this information sheet.

If you would like to be involved, please complete the attached consent form.

Annex II: Consent form

Consent Form

- I have read and understood the information sheet.
- I confirm that the purpose of the study has been explained to me and that I have had the opportunity to ask questions about the research and have had these answered satisfactorily.
- I understand that my participation is voluntary, and that I am free to withdraw at any time without giving any reason.
- I give permission for the researcher to use any photograph(s) and/or videotape(s) made as part of this research project in presentations, publications, for educational uses, or through websites and social media.
- I understand that images on websites can be viewed throughout the world and that some countries may not provide the same level of protection to the rights of individuals as EU/<national> legislation provides.
- I agree that my anonymised research data may be used by others for future research.
- I understand that personal data will be always kept confidential.
- I agree to take part in the study.

----------	----------	----------
Name of respondent	Date	Signature

----------	----------	----------
Name of researcher	Date	Signature

Annex III: Dataset Template

The following template has been developed to specifically gather information on the various AI-MATTERS Dataset

Service Request Identification	Company name	Name of the Company requesting a specific service and specific datasets
	Country	
	Target AI application (if applicable)	
	Target industrial sector	
	Service category requested	
	Node that received the request	
	Status of service request	Initial- Advanced
	Other dataset as needed	
Dataset Overview	Dataset ID	[PartnerName]_[Progressive_Number]_[keyword]
	Partner	
	Provider	
	Short Description	
	Data type and sources	Please leave only one option: • Project Management Data • Financial Data • Customer related Data • Business activity Data • Technical Data • Dissemination & Communication Data • Website and social media analytics Data • Data from previous Projects/other activities internal to the organization • Open dataset • Synthetic data
Dataset Format	Structure description:	Description of the structure and type of the data. (structured | unstructured)

	Type and format	Dataset format, specifying if it is using, for example, CSV, Excel spreadsheet, XML, JSON, etc
	Time coverage	if the dataset has a time dimension, indication of what period does it cover
	Languages:	languages of metadata, attributes, code lists, descriptions.
	Identifiability of data:	reference to identifiability of data and standard identification mechanism
	Naming convention	description about how the dataset can be identified if updated or after a versioning task has been performed, if the dataset is not static.
	Versioning convention	reference to how often is the data updated
	Metadata standards	specification of standards for metadata creation (if any). If there are no standards description of what metadata will be created and how
	Additional comments	indication of other relevant information.
FAIR Dataset	Findability	Will data be identified by a persistent identifier?
		Does the dataset come with some metadata? Are these standards? Which?
		Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?
		Will metadata be offered in such a way that it can be harvested and indexed?
	Accessibility	Where will these dataset be stored? (e.g., Project Repository, Internal Repository…)
		Who is responsible for data storage?
		Who has access to the dataset? (Partner name(s), all partners, open access)
		Data sharing licensing (Proprietary License, Open License, Restricted access etc.)
		How will the data be shared?
		How long will the data remain available and findable?
	Interoperability	What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use?
		In case it is unavoidable that you use uncommon or generate Project specific ontologies or

		vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?
		Will your data include qualified references to other data (e.g. other data from your Project, or datasets from previous research)?
	Reusability	How will you provide documentation needed to validate data analysis and facilitate data re-use? (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)
		Will your data be made freely available in the public domain to permit the widest re-use possible?
Dataset Security	Personal Data	(Y|N): Confirmation about personal data presence in the dataset.
	Anonymized	(Y|N|NA): confirmation if personal data is anonymized
	Secure storage procedures:	Information about how managed data recovery and secure storage was
	Encryption	(Y|N): identification of encryption procedures
	Dataset ethics and legal requirements	describe any Ethics and/or legal requirements relevant to the dataset if not covered in previous questions

Annex IV: Disclaimers for Dissemination & Communication activities

All partners, when communicating and/or disseminating the project, either online or on printed material, should ensure that the following disclaimer and European Commission funding acknowledgement are clearly displayed, alongside the EU emblem, in the following format:

Co-Funded by the European Union, under grant agreement number 101100707

END OF DOCUMENT