Project information	
Full title:	**REmanufacturing, repurposing and recycling energy goods through advanced Mechatronic and Digital technologies**
Acronym:	R3-Mydas
Grant Agreement number:	101138738 (HORIZON-CL4-2023-TWIN-TRANSITION-01-04 – Innovation Action)

Deliverable

D1.2 – Data Management Plan – initial version

Dissemination level:	PU - Public, fully open
Type of deliverable:	DMP - Data Management Plan
Contractual date of delivery:	30 June 2024
Deliverable leader:	NCI
Status - version, date:	Final – v1.0, 2024-06-28
Keywords:	Data Management Plan, R3-Mydas Project, Circular Economy, Remanufacturing, Digital Technologies, Machine Learning, Data Quality Control, Data Storage and Security, FAIR Data Principles, GDPR Compliance

Funded by the European Union

This document is part of a project that is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Find us at www.r3-mydas.eu.

Executive Summary

The R3-Mydas project, focused on remanufacturing, repurposing, and recycling energy goods, utilizes advanced mechatronic and digital technologies to foster sustainable circular value chains. Recognizing the critical role of data in achieving these goals, this Data Management Plan (DMP) outlines a comprehensive strategy for managing the diverse datasets generated and utilized throughout the project.

R3-Mydas seeks to optimize remanufacturing processes across Oil & Gas crankshafts, E-vehicle batteries, and Wind turbines gearboxes. Objectives include reducing programming time by up to 60%, enhancing product quality by 20%, and cutting rework by 30%. The project aims to improve anomaly detection, streamline data handling, and promote high rates of material reuse, thereby reducing lead times and raw material consumption significantly.

This Data Management Plan (DMP) outlines strategies for comprehensive data management throughout the project lifecycle. It covers data collection, storage, access, sharing, and preservation, ensuring integrity, security, and compliance with ethical and legal standards. The plan adheres to FAIR Data Principles, ensuring data is Findable, Accessible, Interoperable, and Reusable (FAIR), and details resource allocation for managing FAIR and non-FAIR data.

The DMP introduces the project by outlining its scope and purpose, providing detailed data summaries, advanced technology integration strategies, and ethical data management frameworks. It ensures GDPR compliance, stakeholder engagement, risk management, and effective monitoring and evaluation. By implementing this DMP, R3-Mydas aims to deliver significant impacts on circular economy practices, ensuring robust data management practices to support its ambitious goals.

Deliverable leader:	Pantelis Lappas (NCI)
Contributors:	All
Reviewers:	Bernhard Peischl (AVL), Dimitra Siaili (ITML)
Approved by:	Athanasios Poulakidas, Anastasia Panitsa (NCI)

Document History			
Version	**Date**	**Contributor(s)**	**Description**
0.1	2024-06-21	P. Lappas, A. Poulakidas	Complete draft, shared for internal review
0.2	2024-06-27	B. Peischl, M. Lebenbauer	Reviewed by AVL and saved as a new version
0.3	2024-06-27	P. Lappas	Adapted by NCI considering reviewer's comments (both ITML & AVL). Prepared final version.
1.0	2024-06-28	A. Poulakidas, A. Panitsa	Version to be submitted after final QA

Table of Contents

List of Tables

Definitions, Acronyms and Abbreviations

Acronym/ Abbreviation	Title
AE	Affiliated Entity
AI	Artificial Intelligence
API	Application Programming Interface
AP	Associated Partner
COO	Coordinator
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoperable, Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
IDS	International Data Spaces
ISO	International Organization for Standardization
KPI	Key Performance Indicator
ML	Machine Learning
Re-Mydas	REmanufacturing, repurposing and recycling energy goods through advanced Mechatronic and Digital technologies
SSH	Social Sciences and Humanities
WP	Work Package

1 Introduction

1.1 Background Information about the Project

The R3-Mydas project puts a special emphasis on remanufacturing, repurposing, and recycling energy goods by employing advanced mechatronic and digital technologies to support sustainable circular value chains. While remanufacturing is well-regarded as an effective strategy for moving towards a circular economy, there remains a critical need for enhanced research and experimental observations. These improvements are necessary to boost the traceability and reliability of final products from the end-user perspective. Furthermore, there is a pressing need for more effective impact monitoring methods.

The primary objective of R3-Mydas is to develop a multi-actor framework that integrates innovative digital technologies and advanced mechatronics with newly developed approaches from social sciences and humanities (SSH). This framework aims to achieve functional, environmental, and economic sustainability in the remanufacturing of energy goods at the factory level. The project targets three demonstrators: Oil & Gas crankshafts, E-vehicles batteries, and Wind turbines gearboxes.

To achieve these objectives, R3-Mydas incorporates various cutting-edge technologies and methodologies. These include machine learning for process and quality control, marketplace development, graph models for defects detection, and digital twins. Advanced mechatronics such as additive manufacturing, laser-cladding, and automated disassembly/reassembly are also key components. Additionally, the project uses advanced acceptance models and a legal and ethical framework to ensure compliance and acceptance.

The expected outcomes of the project are substantial. For Oil & Gas crankshafts, there is an anticipated reduction in programming time for remanufacturing by up to 60%, an increase in product quality by up to 20%, and a reduction in rework by up to 30%. For E-vehicles batteries, the project aims to improve detection of tiny deviations from normal behaviour by up to 30%, achieve 50% faster anomaly localization, increase the number of different modality data streams handled by up to 30%, and speed up the fusion process by up to 20%. For Wind turbines gearboxes, the targets include achieving a reuse rate of up to 99%, a prevention rate of 90%, a reduction in lead time by 75%, and potential raw material savings of up to 85%.

The project will also deliver a marketplace that associates each remanufactured product or service/component for remanufacturing with a Digital Passport-like set of information, ensuring full traceability. Additionally, a dedicated training program will be designed and delivered by EITM, targeting the project's remanufacturing value chains with over 100 training hours and engaging more than 100 diverse stakeholders.

The consortium consists of 15 partners, with NCI Company serving as the coordinator. The project comprises nine work packages (WPs), each with specific objectives and deliverables. WP1 (5 deliverables) focuses on strategic project management, day-to-day coordination, communication, quality and risk management, and the data management plan (DMP). WP2 (3 deliverables) aims to develop and validate a circular value chain for remanufacturing Oil & Gas crankshaft components, incorporating a digital-based design methodology, laser-cladding path planning, processing optimization, and method validation. WP3 (3 deliverables) targets the remanufacturing of E-vehicles batteries through purification and reuse of raw materials, integration of mechatronics with digital systems, development of an IoT framework for monitoring, and anomaly detection.

WP4 (5 deliverables) is dedicated to the remanufacturing of Wind turbines gearboxes, focusing on new materials and design methods, options for remanufacturing, and cost simulation. WP5 (7 deliverables) addresses standardization and impact evaluation, aiming to create a framework for standardization, evaluate environmental and socio-economic impacts, and identify new value chains. WP6 (4 deliverables) involves the development of cognitive robotics and quality control systems, incorporating monitoring modules, data fusion, process state estimation, machine learning for quality control, and a digital marketplace.

WP7 (3 deliverables) focuses on dissemination and communication, developing plans for communication and dissemination, engaging with stakeholders, liaising with standards and other projects, and promoting post-project commercialization. WP8 (6 deliverables) aims to develop a viable strategy for exploiting the R3-Mydas outcomes, including defining business models, exploitation plans, market analysis, IPR strategy, and training for adoption. Finally, WP9 (1 deliverable) ensures compliance with ethics requirements, safeguarding the project's adherence to ethical standards.

Through these work packages, R3-Mydas aims to deliver significant impacts across the targeted value chains, enhancing efficiency, product quality, and sustainability, ultimately contributing to the broader goals of a circular economy and sustainable industrial practices.

Table 1: The R3-Mydas Consortium.

Number[1]	Name	Country	Short name
1(CO)	NETCOMPANY-INTRASOFT SA	Luxemburg	NCI
2	EUROPEAN FEDERATION FOR WELDING JOINING AND CUTTING	Belgium	EWF
3	EIT MANUFACTURING SOUTH SRL	Italy	EITM
4	FLENDER FINLAND OY	Finland	FLE-FI
4.1(AE)	FLENDER GMBH	Germany	FLE
5	AVL LIST GMBH	Austria	AVL

[1] CO: Coordinator. AE: Affiliated Entity. AP: Associated Partner.

Number[1]	Name	Country	Short name
6	TALLERES MECANICOS COMAS SLU	Spain	TMCOMAS
7	SPIN ROBOTICS IVS	Denmark	SPIN
8	ASOCIATION DE INVESTIGACION METALURGICA DEL NOROESTE	Spain	AIMEN
9	LAPPEENRANNAN-LAHDEN TEKNILLINEN YLIOPISTO LUT	Finland	LUT
10	INFORMATION TECHNOLOGY FOR MARKET LEADERSHIP	Greece	ITML
11	DEEP BLUE SRL	Italy	DBL
12	CHAROKOPEIO PANEPISTIMIO	Greece	HUA
13	IKERLAN S. COOP	Spain	Ikerlan
14	ZIKNES TECHNOLOGY SL	Spain	Ziknes
15 (AP)	CSEM CENTRE SUISSE D'ELECTRONIQUE ET DE MICROTECHNIQUE SA - RECHERCHE ET DEVELOPPEMENT	Switzerland	CSEM

1.2 Purpose and Scope of the Data Management Plan

This deliverable aims to compile all the data managed by consortium partners throughout the R3-Mydas project. Notably, the Data Management Plan (DMP) will be continuously updated and maintained as online live pages within the project's Data Management System (DMS) for the duration of the project.

1.3 Document Structure

The Data Management Plan (DMP) is structured to provide a comprehensive framework for managing data throughout the R3-Mydas project lifecycle. It is organized into several key chapter to ensure clarity and coherence in handling project data:

- **Introduction:** This chapter provides an overview of the R3-Mydas project, detailing its background, scope, and objectives. It explains the importance of effective data management within the project and outlines the structure of the DMP. The roles and responsibilities of key stakeholders involved in the data management process are also defined to ensure clarity and accountability throughout the project's lifecycle.
- **Data Summary:** This chapter offers a comprehensive summary of the types of data generated and utilized within R3-Mydas. It includes data sources, collection methods, and procedures, along with formats, volume, and velocity of data. Storage and security protocols are described to safeguard data integrity. Policies for data access, sharing, quality control, retention, and destruction are detailed, ensuring proper data handling from collection to disposal.
- **Advanced Technology and Machine Learning Integration:** This part discusses the integration of advanced technologies and machine learning models into the

project. It highlights how innovative digital solutions could be utilized to enhance remanufacturing processes. In addition to this, it provides an overview of machine learning algorithms and techniques employed in the project, including data preprocessing, model training and validation, deployment strategies, and ethical considerations in model development.

- **FAIR Data:** The FAIR Data chapter outlines strategies to ensure that data generated by R3-Mydas is Findable, Accessible, Interoperable, and Reusable. It details procedures for making data easily discoverable and accessible, ensuring compatibility with various systems and enabling data reuse across different contexts. This approach maximizes the usability of data and aligns with open data standards.
- **Allocation of Resources**: Details the costs associated with managing FAIR and non-FAIR data, responsibilities for data management within the consortium, and budgeting for long-term data preservation.
- **Ethical Data Management and Compliance Strategy:** Focused on ensuring ethical and legal compliance, this chapter details how R3-Mydas adheres to the General Data Protection Regulation (GDPR) and other relevant data protection policies. It includes guidelines for personal data processing, data protection roles, consent procedures, rights of data subjects, and measures for data breach response. The chapter emphasizes the importance of data governance, compliance, and risk management.
- **Conclusion**: Summarizes key insights and recommendations derived from the DMP, highlighting next steps for implementing effective data management practices within the R3-Mydas project.
- **Appendix A**: The appendix outlines a comprehensive data survey conducted with R3-Mydas project partners via an Excel file to gather essential information on dataset management.
- **Annex 1:** Provides a consent form template to be used in R3-Mydas when needed. The form can be adapted to consent to a variety of activities.

1.4 Key Stakeholders and their Roles and Responsibilities

The R3-Mydas project involves a consortium of diverse stakeholders, each playing critical roles in advancing the objectives of remanufacturing energy goods through advanced mechatronic and digital technologies. These stakeholders include partners from various sectors, each bringing unique expertise and resources to the project. Their collaborative efforts ensure comprehensive data management, fostering transparency, efficiency, and impactful outcomes across the project's lifecycle.

- **NCI** leads the project as the coordinator, ensuring seamless communication, effective project management, and successful achievement of the project's

goals. NCI also spearheads the development of the R3-Mydas marketplace, an innovative platform facilitating the buying and selling of remanufactured goods, components, and services while ensuring product traceability and reliability.

- **EWF** focuses on identifying standardization gaps and ensuring the development of standardization documents by relevant bodies. Their role is crucial in maintaining high standards and promoting the adoption of remanufacturing technologies.
- **EITM** aims to expand its community and training opportunities, leveraging its participation in R3-Mydas. EITM plays a pivotal role in devising and implementing methodologies for developing new training curricula and business models in remanufacturing.
- **CSEM** contributes significantly with its expertise in cognitive robotics and quality control modules. Their involvement ensures the integration of cutting-edge technologies for monitoring and optimizing remanufacturing processes.
- **FLE-FI** brings their extensive experience in the service market for wind turbine gearboxes. They lead the demo case for wind turbines, using lessons learned to enhance customer marketing and update product design philosophies, ultimately contributing to more efficient future applications.
- **AVL** specializes in engineering services for the automotive industry and leads the demo case for E-vehicles. They focus on developing a generic methodology for graph-neural network-based anomaly detection, helping quality engineers apply these techniques in their specific settings.
- **TMCOMAS** leads the demo case for Oil & Gas components, utilizing advanced digital-based procedures for remanufacturing. Their sales department will exploit the developed repair procedure as part of a new update for the cladding cell.
- **SPIN** focuses on commercializing new software for advanced anomaly detection, integrating it into the company's product portfolio. Their expertise in disassembly robots is crucial for the demo case involving E-vehicles batteries.
- **AIMEN** leads the research on laser-cladding technologies for the Oil & Gas demo case. They also provide Life Cycle Assessment (LCA), Life Cycle Costing (LCC), and Social Life Cycle Assessment (s-LCA) methodologies for evaluating new remanufacturing processes at low TRL levels.
- **LUT** leads research for the wind turbine gearbox demo case. They develop simulation methods and digital twins, assessing and validating the impact on industrial logistics and business models.
- **ITML** contributes big data analytics and advanced AI services for EV batteries, ensuring effective assessment and testing of components during disassembly and reassembly.
- **DBL** focuses on the development and validation of assessment methods for ethics and liability aspects. They also work on creating skills and training sets for future manufacturing environments.

- **HUA** develops Technology Acceptance Models (TAM) and Unified Theory of Acceptance and Use of Technology (UTAUT) models, ensuring the technological acceptance of remanufacturing approaches by stakeholders.
- **IKERLAN** develops structural reliability models for helical geometries and improved induction hardening simulation tools, contributing to the wind turbine gearbox demo case.
- **Ziknes** creates a software platform for advanced toolpaths in robotics, reducing operator intervention and enhancing process control for new remanufacturing processes.

These stakeholders play integral roles in the R3-Mydas project, each contributing unique expertise and resources to advance the goals of remanufacturing energy goods through innovative technologies and sustainable practices. Their collaborative efforts ensure comprehensive data management, fostering transparency, efficiency, and impactful outcomes across the project's lifecycle.

2 Data Summary

2.1 Overview of Data Sources and Types

This section provides an overview of the datasets used in the R3-Mydas project, including their names, descriptions, and the partners responsible for each dataset. It also details the roles and responsibilities of each partner, the tasks and work packages associated with the data, the sources from which the data will be collected, and the types of data involved.

Table 2: Overview of Data Sources and Types.

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
Battery ageing lab data	Battery cell measurements for the state of health (SOH)	AVL	Clear responsibilities defined within AVL.	T3.3	WP3	battery testbed	time-series data; meta-data on the meaning of the data
Fleet data from vehicles	Vehicle sensor data.	OEM-s; Tier-1-s; fleet operators	Clear responsibilities at AVL and customer side.	T3.3	WP3	Fleet operators	time-series data; meta-data on the meaning of the data
Tasks descriptions	Study human-machine interaction, envision future scenarios.	DBL, TMCOMAS, AVL, FLENDER	DBL	T2.1, T3.1, T4.1, T5.4, T5.6	WP2, WP3, WP4, WP5	Informed consents, surveys, interviews, observations. Consent data segregated; research data anonymized.	Qualitative, structured data

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
Workstation analysis	Study human-machine interaction, envision future scenarios.	DBL, TMCOMAS, AVL, FLENDER	DBL	T2.1, T3.1, T4.1, T5.4, T5.6	WP2, WP3, WP4, WP5	Informed consents, surveys, interviews, observations. Consent data segregated; research data anonymized.	Qualitative, structured data
Users profiles	Assess user skills, define training plan.	DBL, TMCOMAS, AVL, FLENDER	DBL	T2.1, T3.1, T4.1, T5.4, T5.6, T8.1	WP2, WP3, WP4, WP5, WP8	Informed consents, surveys, interviews, observations. Consent data segregated; research data anonymized	Qualitative, structured data
Ergonomics data	To understand the users' needs and support the to-be scenario definition	DBL, TMCOMAS, AVL, FLENDER	DBL	T2.1, T3.1, T4.1, T5.4, T5.6	WP2, WP3, WP4, WP5	Informed consents, surveys, interviews, observations. Consent data segregated; research data anonymized	Qualitative structured data and quantitative unstructured data: workload (physical and cognitive), trust, stress

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
Neurophysiological data (eye tracking)	Understand user needs, support scenario definition and evaluation.	DBL, TMCOMAS, AVL, FLENDER	DBL	T2.1, T3.1, T4.1, T5.4, T5.6	WP2, WP3, WP4, WP5	Informed consents, surveys, interviews, observations. Consent data segregated; research data anonymized	Quantitative data from wearable devices measuring stress, workload, and arousal.
Neurophysiological data (EEG)	Understand user needs, support scenario definition and evaluation.	DBL, TMCOMAS, AVL, FLENDER	DBL	T2.1, T3.1, T4.1, T5.4, T5.6	WP2, WP3, WP4, WP5	Informed consents, surveys, interviews, observations. Consent data segregated; research data anonymized	Quantitative data from wearable devices measuring stress, workload, and arousal.
Neurophysiological data (galvanic)	Understand user needs, support scenario definition and evaluation.	DBL, TMCOMAS, AVL, FLENDER	DBL	T2.1, T3.1, T4.1, T5.4, T5.6	WP2, WP3, WP4, WP5	Informed consents, surveys, interviews, observations. Consent data segregated; research data anonymized	Quantitative data from wearable devices measuring stress, workload, and arousal.

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
Microsoft Office	Storing EITM-related materials, contributions, and activities in an internal SharePoint for the project.	EITM Internal access	SharePoint managed by IT Admin, accessed by EITM staff. Senior and Program Managers update data.	Task 8.1 Task 8.2 Task 8.3 Task 8.4	WP8 WP1	Survey, Interviews, desk research from primary and secondary analysis, internal resources of the EITM (platforms and networks)	All different types of data in structured format.
Microsoft Form	Gathering data on use case development for new business models in WP8.	ALL	SharePoint managed by IT Admin, accessed by EITM staff. Senior and Program Managers update data.	Task 8.1 and 8.2	WP8	Survey designed for the project use	Data from survey will be in both qualitative and quantitative forms and structured
Standards	Compiling standards for a recommendation document to ISO/CEN.	All	EWF	5.1	5	excel files, shared documents to be filled and meetings	Excels, docs; Both qualitative and quantitative; Structured
Communication and dissemination material	Information on dissemination events, project materials (flyers, posters, roll-ups), event images, and more.	All	EWF	7.1	7	forms, shared documents to be filled and meetings	Excels, docs; Qualitative; Structured

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
Stakeholder Analysis and community building strategy	Information on the stakeholders and community building.	All	EWF	7.2; 7.3; 7.4	7	forms, shared documents to be filled and meetings	Excels, docs; Qualitative; Structured
Gear tester data	Excel data	FLE FI	Research engineer	4,3	WP4	Experimental data (Software and measurements)	Qualitative and structured
Journal bearing tester data	Excel data	FLE FI	Research engineer	4,4	WP4	Experimental data (Software and measurements)	Qualitative and structured
Flange connection tester data	Excel data	FLE FI	Research engineer	4,2	WP4	Experimental data (Software and measurements)	Qualitative and structured
FEA simulations	Reports	FLE FI	Research engineer	4,2; 4,3; 4,4	WP4	Experimental data (Software and measurements)	Qualitative and structured
Time-domain simulations	Reports	FLE FI	Research engineer	4,5	WP4	Experimental data (Software and measurements)	Qualitative and structured
Test pictures	Reports	FLE FI	Research engineer	4,2; 4,3; 4,4	WP4	Phone pictures	Qualitative and structured
Responses to user acceptance questionnaires	Help investigate the consumer acceptance of R3-Mydas' technologies	HUA	HUA (collection and processing of data)	T5.3	WP5	Questionnaires (and potentially interviews)	Qualitative & structured

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work pack age no.	Data source	Data type
Residual stresses for induction hardened gear	To know the post-manufacturing state of the gear	Ikerlan	Mikel Escalero, T4.3 leader	T4.3	WP4	Calculations performed by Ansys and in-house tools	Quantitative, solved finite element model
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	EV dataset for battery aging models, aiding health estimation and diagnostics.	ITML	• Advise ITML and its employees on Data Protection regulations. • Ensure policy alignment with Data Protection laws. • Guide on Data Protection Impact Assessments (DPIAs). • Liaise with Data Protection Authorities (DPAs). • Manage notifications to DPAs for Personal Data processing. • Establish a system for handling Data Subject requests promptly. • Inform senior management of potential	Task 3.1, Task 3.3, Task 3.4, Task 3.5	WP3	1. Proprietary (synthetic) datasets provided by the pilot leader 2. Benchmarked data (publicly available)	Quantitative data will be utilized. Regarding the structure of the data, this is not yet determined.
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	Battery dataset analyses charging patterns, detects anomalies, monitors health.	ITML		Task 3.1, Task 3.3, Task 3.4, Task 3.5	WP3		
SICWELL Dataset	This dataset models and diagnoses lithium-ion batteries in electric vehicles.	ITML		Task 3.1, Task 3.3, Task 3.4, Task 3.5	WP3		

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
			legal consequences for Data Protection violations.				
Minutes of the meetings	Minutes of the meetings, synchronization of actions between team members.	All	Work package leaders			Summaries of meetings	Text, pictures, tables, figures.
Gearbox simulation models	Simulation models including input and output data	LUT, Flender	Will be selected based on use case		WP4	LUT	Structured
Gearbox models	CAD and topological models	LUT, Flender	Will be selected based on use case		WP4	Flender	Structured
Measurement data	Measurements from gearbox performed by Flender	LUT, Flender	Will be selected based on use case		WP4	Flender	Structured
Experiment protocols	Documentation of how experiments are conducted and what are desired outputs	LUT, Flender	Will be selected based on use case		WP4	Flender, LUT	Structured
Scientific papers in production	Manuscript and associated attachments	LUT, Flender	Will be selected based on use case		WP4	LUT	Structured

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
Impact of remanufacturing process on industrial logistics	Data insights, benchmarks, methodologies, market analysis, forecasting.	Not available at present (LUT Business School & external research agency(?))	LUT oversees data collection, processing, and storage. External agency handles data collection, processing, and presentation according to standards.	Task 5.5	WP 5	Survey and/or interviews	* Survey - quantitative/qualitative; structured * Interviews - qualitative; semi-structured
Real crankshaft dimensions (geometric data)	These will affect the programming of the robot performing the part scanning for repair.	TMCOMAS	Technical department using the Mydas solution.	N/A	WP2, task 2.1, subtask 2.1.1	customer drawings or physical measurements on real crankshaft	Geometrical dimensions. Quantitative. Unstructured
Robot program for scanning the part	This program will actuate the ABB robot to perform the initial scanning of the part	AIMEN/TMCOMAS	Technical department using the Mydas solution.	N/A	WP2, task 2.1, subtask 2.1.1	generated by robot programming software	N/A
cloud point data (scanning the crankshaft)	Data used to generate robot program for cladding process execution.	AIMEN/TMCOMAS	Technical department using the Mydas solution.	N/A	WP2, task 2.1, subtask 2.1.1	generated by the 3D scanner or snapshot camera during crankshaft scanning	topographical information about the part surface and dimension
laser cladding parameters	These data are needed to actuate the cladding cell to perform good welding process	AIMEN/TMCOMAS	Technical department using the Mydas solution.	N/A	WP2, task 2.2	generated by metallurgical trials during task 2.2	laser power, powder feed rate, welding speed, cladding head tilting, etc.

Name of the dataset	Dataset description	Partner name	Responsi-bilities and roles	Task no.	Work package no.	Data source	Data type
Robot program to weld the part	This program will actuate the ABB robot to perform the initial scanning of the part	AIMEN/ZIKNES	Technical department using the Mydas solution.	N/A	WP2, task 2.1, subtask 2.1.1	generated by robot programming software	N/A
welding quality monitoring data	These data will be used to assess the quality of the cladding through the process	AIMEN/TMCOMAS	Technical department using the Mydas solution.	N/A	WP2, task 2.2	generated by some sensor (IR camera, pyrometer or thermographic camera)	numerical values to be correlated or analysed by a dedicated software
Robot program for scanning the part after the welding	This program will actuate the ABB robot to perform the final scanning of the part	AIMEN/TMCOMAS/ZIKNES	Technical department using the Mydas solution.	N/A	WP2, task 2.3	generated by robot programming software	N/A
final validation of the MYDAS solution	Data concerning NDT (non-destructive test) and geometrical dimensions	TMCOMAS	Technical department using the Mydas solution.	N/A	WP2, task 2.3	generated by manual measuring processes and NDT inspection processes	Geometrical dimensions. NDT reports

Table 2 contains a detailed overview of various data sources and types, categorized by their applications, responsible partners, and other pertinent attributes within a research project. The data sources are diverse, encompassing both qualitative and quantitative data, and serve multiple research objectives, including battery ageing, vehicle fleet data, human-machine interaction studies, ergonomic assessments, neurophysiological measurements, user acceptance, and more. Here is a summary of the main insights.

2.1.1 Battery Data

Battery data consists of:

- Battery Ageing Lab Data: Collected by AVL, focuses on battery cell measurements for the state of health (SOH), involving time-series data and meta-data. It falls under Task 3.3 and WP3.

- Fleet Data from Vehicles: Gathered by OEMs, Tier-1s, and fleet operators, this data pertains to vehicle sensor measurements. Responsibilities are shared between AVL and customers, also under Task 3.3 and WP3.
- Lithium-ion Battery Aging Dataset: Managed by ITML, this dataset aids in health estimation and diagnostics of batteries using both proprietary and benchmarked data, associated with multiple tasks in WP3.

2.1.2 Human-Machine Interaction and Ergonomics

Human-machine interaction and ergonomics data are groups as follows:

- Tasks Descriptions, Workstations Analysis, Users Profiles, Ergonomics Data: All managed by DBL, TMCOMAS, AVL, and FLENDER, these datasets study various aspects of human-machine interaction and user skills, involving informed consents, surveys, interviews, and observations. The data types are both qualitative and structured.
- Neurophysiological Data (Eye Tracking, EEG, Galvanic): Collected by the same partners, these datasets use wearable devices to measure stress, workload, and arousal, providing quantitative data for scenario definition and evaluation.

2.1.3 Administrative and Project Management Data

The data under this groups are the following:

- Microsoft Office and Form Data: Used for storing project materials and gathering data on use case development, managed by EITM with access controlled by IT Admins. Data is structured and qualitative/quantitative.
- Standards, Communication and Dissemination Material, Stakeholder Analysis: Managed by EWF, these involve compiling standards, dissemination event data, and stakeholder information. The data is both qualitative and quantitative, structured in formats like Excel and documents.

2.1.4 Experimental and Simulation Data

Experimental and simulation data are groups as follows:

- Gear Tester, Journal Bearing Tester, Flange Connection Tester, FEA Simulations, Time-domain Simulations, Test Pictures: Managed by FLE FI, these datasets involve experimental data, software measurements, and simulation reports, focusing on structured and qualitative data.
- Real Crankshaft Dimensions, Robot Programs, Cloud Point Data, Laser Cladding Parameters, Welding Quality Monitoring Data, Final Validation: These datasets, related to the MYDAS solution, are managed by AIMEN, TMCOMAS, and ZIKNES, involving quantitative and unstructured data for robotic and welding processes.

2.1.5 User Acceptance and Market Analysis Data

The following groups of data were recognized:

- Responses to User Acceptance Questionnaires: Collected by HUA, this data helps investigate consumer acceptance of new technologies, involving structured qualitative data.
- Impact of Remanufacturing Process on Industrial Logistics: Managed by LUT and potentially an external research agency, this dataset includes data insights, benchmarks, and market analysis, collected through surveys and interviews.

2.1.6 Miscellaneous Data

Miscellaneous data contains:

- Minutes of Meetings: Summarizes meetings and actions among team members, involving text, pictures, tables, and figures.
- Scientific Papers in Production: Includes manuscripts and associated attachments, managed by LUT and Flender, in a structured format.

Overall, the project encompasses a wide range of data types and sources, each tailored to specific research needs, with clear responsibilities and structured methodologies for data collection, processing, and analysis.

2.2 Data Collection Methods and Procedures

This section outlines the methods and procedures for data collection within the project. It includes the frequency of data collection, the availability of the data, and the specific methodologies employed to gather data from various sources (see Table 3).

Table 3: Data Collection Methods and Procedures.

Name of the dataset	Partner name	Data collection frequency	Data availability
Battery ageing lab data	AVL	data on historic cell ageing tests	now
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	continuous data recording	now
Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	Iterative approach with ongoing data collection during activities involving users and partners.	M12
Workstations analysis	DBL, TMCOMAS, AVL, FLENDER	Iterative approach with ongoing data collection during activities involving users and partners.	M12
Users profiles	DBL, TMCOMAS, AVL, FLENDER	Iterative approach with ongoing data collection during activities involving users and partners.	M12

Name of the dataset	Partner name	Data collection frequency	Data availability
Ergonomics data	DBL, TMCOMAS, AVL, FLENDER	Iterative approach with ongoing data collection during activities involving users and partners.	M12
Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER	Specific experiments will be organized at specific moments during the project activities	M12
Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER	Specific experiments will be organized at specific moments during the project activities	M12
Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER	Specific experiments will be organized at specific moments during the project activities	M12
Microsoft Office	EITM Internal access	ongoing	M12
Microsoft Form	ALL	ongoing until M12	The first release will be available by June.
Standards	All	On an ongoing basis	M6
Communication and dissemination material	All	On an ongoing basis	
Stakeholder Analysis and community building strategy	All	On an ongoing basis	M12
Gear tester data	FLE FI	weekly	Q2/2024
Journal bearing tester data	FLE FI	weekly	Q2/2024
Flange connection tester data	FLE FI	weekly	Q2/2024
FEA simulations	FLE FI	weekly	Q2/2024
Time-domain simulations	FLE FI	weekly	Q2/2024
Test pictures	FLE FI	weekly	Q2/2024
Responses to user acceptance questionnaires	HUA	At specific intervals	M18
Residual stresses for induction hardened gear	Ikerlan	N/A	End of first year
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	Data collection frequency will be defined together with the pilot leader to meet the requirements of Demo Case 2.	will be defined by the pilot leader when the task starts
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML		
SICWELL Dataset	ITML		
Minutes of the meetings	All	Weekly	At the start
Gearbox simulation models	LUT, Flender		After half a year
Gearbox models	LUT, Flender		4 months from project start

Name of the dataset	Partner name	Data collection frequency	Data availability
Measurement data	LUT, Flender		1 year from project start
Experiment protocols	LUT, Flender		8 months from project start
Scientific papers in production	LUT, Flender		5 months from project start
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	To be determined (Survey - fixed period e.g., spring/summer 2025; Interviews - ongoing 2025)	Year 3 (2025)
Real crankshaft dimensions (geometric data)	TMCOMAS	On-going basis. At the begging of the repair process.	From M1
Robot program for scanning the part	AIMEN/TMCOMAS	On-going basis. At the begging of the repair process.	M6
cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS	On-going basis. At the begging of the repair process.	from M6
laser cladding parameters	AIMEN/TMCOMAS	On-going basis.	M6-M20
Robot program to weld the part	AIMEN/ZIKNES	On-going basis. At the begging of the repair process.	from M6 approx.
welding quality monitoring data	AIMEN/TMCOMAS	On-going basis.	M6-M20 approx.
Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES	On-going basis. At the begging of the repair process.	from M6 approx.
final validation of the MYDAS solution	TMCOMAS	on-going basis	M18-M36 approx.

The data collection methods across various datasets in the project exhibit a structured approach tailored to specific needs and timelines. Beginning with foundational datasets such as Battery Ageing Lab Data managed by AVL, historical data on cell ageing tests are currently underway, ensuring a comprehensive understanding of battery performance over time. Similarly, Fleet Data from Vehicles, sourced from OEMs, Tier-1 suppliers, and fleet operators, undergoes continuous recording, providing real-time insights into vehicle performance and usage patterns.

Human-machine interaction and ergonomic datasets managed by DBL, TMCOMAS, AVL, and FLENDER follow an iterative approach. This includes Tasks Descriptions, Workstations Analysis, Users Profiles, and Ergonomics Data, with ongoing collection aligned with project activities involving users and partners. Scheduled until Month 12 (M12), this approach ensures that evolving user needs, and workplace dynamics are captured effectively throughout the project lifecycle.

Neurophysiological data, encompassing Eye Tracking, EEG, and Galvanic Data, is collected through specific experiments organized at key project milestones. This targeted approach, implemented by DBL, TMCOMAS, AVL, and FLENDER, allows for

detailed insights into user responses and physiological states, contributing crucial data for scenario definition and evaluation by Month 12 (M12).

In terms of project management tools, Microsoft Office and Microsoft Forms serve as essential repositories for project-related materials and user feedback. Managed internally and accessible to project staff by Month 12 (M12), these tools facilitate streamlined data management and stakeholder engagement.

Technical testing datasets, such as those from Gear Tester, Journal Bearing Tester, Flange Connection Tester, FEA Simulations, Time-Domain Simulations, and Test Pictures managed by FLE-FI, are collected weekly starting from the second quarter of 2024 (Q2/2024). This frequent data collection ensures that experimental results and test outcomes are promptly integrated into ongoing project activities.

The compilation of Standards and Communication and Dissemination Materials by all partners occurs continuously throughout the project, supporting regulatory compliance and effective project visibility without a predefined endpoint.

The Impact of Remanufacturing Process on Industrial Logistics, managed by LUT Business School and an external research agency, involves periodic surveys and ongoing interviews until 2025 (Year 3). This longitudinal approach ensures comprehensive data collection for assessing industrial impacts and logistic efficiencies.

Finally, Robotic and Repair Process Data, including Real Crankshaft Dimensions, Robot Programs, Cloud Point Data, Laser Cladding Parameters, Welding Quality Monitoring Data, and Final Validation of the R3-Mydas Solution, managed by AIMEN, TMCOMAS, and ZIKNES, are collected continuously starting from Month 6 (M6) and extend to Month 36 (M36). This extensive timeframe ensures thorough validation and optimization of robotic processes and repair methodologies throughout the project lifecycle.

In conclusion, the project's approach to data collection exemplifies a harmonious blend of structured methodologies and iterative processes. This dual approach not only guarantees the continual relevance and adaptability needed to meet evolving project requirements and technological advancements but also ensures that data availability aligns closely with specific timelines. The emphasis on focused experimentation, particularly evident in the acquisition of neurophysiological data through targeted experiments, underscores the project's commitment to refining user-centric designs and operational strategies. Moreover, the seamless integration of continuous project support mechanisms, including robust management tools, standards compilation, and ongoing stakeholder engagement, reinforces real-time oversight and collaboration. Lastly, the longitudinal assessment of industrial and robotic process data underscores the project's dedication to informed decision-making and strategic planning, ultimately fostering innovation and efficiency across all project domains.

2.3 Data Formats, Volume, and Velocity

This section describes the formats in which data will be stored, the expected volume of data to be managed, and the velocity at which data will be generated and processed (see Table 4). These details are crucial for understanding the technical requirements and challenges associated with data handling.

Table 4: Data Formats, Volume, and Velocity.

Name of the dataset	Partner name	Data format	Data volume	Data velocity
Battery ageing lab data	AVL	mf4	GB	1 MB/h per cell
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	mf4	GB	50 MB/h per car
Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	Excel and graphic representations	NA	NA
Workstations analysis	DBL, TMCOMAS, AVL, FLENDER	Excel and graphic representations, CAD files	NA	NA
User profiles	DBL, TMCOMAS, AVL, FLENDER	Excel and graphic representations	NA	NA
Ergonomics data	DBL, TMCOMAS, AVL, FLENDER	Excel and graphic representations	NA	NA
Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER	Excel and CSV files	GBs	NA
Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER	Excel and CSV files	GBs	NA
Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER	Excel and CSV files	GBs	NA
Microsoft Office	EITM Internal access	SharePoint database	55 MB	NA
Microsoft Form	ALL	Microsoft Form database and Excel file	20-30 KB	NA
Standards	All	excel, docs, pdf	MB	slow KB/month
Communication and dissemination material	All	excel, docs, pdf	GB	

Name of the dataset	Partner name	Data format	Data volume	Data velocity
Stakeholder Analysis and community building strategy	All	excel, docs, pdf	MB	a few KB/month
Gear tester data	FLE FI	Excel	12 Gb	2 Gb /month
Journal bearing tester data	FLE FI	Excel	12 Gb	2 Gb /month
Flange connection tester data	FLE FI	Excel	12 Gb	2 Gb /month
FEA simulations	FLE FI	Excel	40 Gb	5 Gb /Month
Time-domain simulations	FLE FI	Excel	10 Gb	500 Mb /month
Test pictures	FLE FI	Excel	10 Gb	100 Mb /month
Responses to user acceptance questionnaires	HUA	Excel	<1MB	<1MB per year
Residual stresses for induction hardened gear	Ikerlan	.wbpj extension	150 GB each	Total 150 GB
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	Possible data formats that will be used are: csv, hdf5, Excel, pickle.	in the range of 1-50GB	meet the requirements of Demo Case 2.
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML			
SICWELL Dataset	ITML			
Minutes of the meetings	All	Word, Power point	max 2GB	200MB/week
Gearbox simulation models	LUT, Flender	XML, binary, csv, excel, MATLAB, python scripts	10GB max	
Gearbox models	LUT, Flender	CAD formats, power point	4GB max	
Measurement data	LUT, Flender	Excell	200MB	
Experiment protocols	LUT, Flender	Excell, Word	100MB	
Scientific papers in production	LUT, Flender	Word, text, LaTeX	100MB	
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	* Survey - CSV & Excel; * Interviews Word & Excel	1 MB	
Real crankshaft dimensions (geometric data)	TMCOMAS	excel file. Word file, pdf or dwg	up to 2-3 Mb	N/A
Robot program for scanning the part	AIMEN/TMCOMAS	.pgf (ABB file format)	up to 1Mb	N/A
cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS	.stl (cloud point file)	50-200 MB	N/A

Name of the dataset	Partner name	Data format	Data volume	Data velocity
laser cladding parameters	AIMEN/TMCOMAS	excel and/or word file	up to 2-3 Mb	N/A
Robot program to weld the part	AIMEN/ZIKNES	.pgf (ABB file format)	up to 1Mb	N/A
welding quality monitoring data	AIMEN/TMCOMAS	.tiff	to be defined	N/A
Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES	.pgf (ABB file format)	up to 1Mb	N/A
final validation of the MYDAS solution	TMCOMAS	pdf	up to 2-3 Mb	N/A

These datasets highlight a structured approach using Excel and graphic representations for tasks' descriptions, workstations analysis, user profiles, and ergonomics data, ensuring compatibility and ease of integration across project phases.

The dataset battery ageing lab data managed by AVL is stored in the mf4 format, generating approximately 1 MB of data per hour per cell. Similarly, fleet data from vehicles, sourced from OEMs, Tier-1 suppliers, and fleet operators, is formatted in mf4, accumulating data at a rate of 50 MB per hour per car.

Neurophysiological data acquisition, including eye tracking, EEG, and galvanic data, utilizes Excel and CSV files, capturing data volumes in gigabytes (GBs), reflecting the project's meticulous focus on detailed physiological insights. Operational data like Microsoft Office applications are managed through SharePoint databases, handling up to 55 MB of data. Conversely, user-generated data via Microsoft Forms is lightweight, typically between 20-30 KB per entry, demonstrating efficient data handling for feedback and assessment purposes.

Technical testing and simulations conducted by FLE-FI involve substantial data volumes and moderate velocities, with FEA simulations requiring up to 40 GB of storage and generating 5 GB of new data per month. Similarly, gearbox simulation models managed by LUT and Flender encompass diverse formats such as XML, binary, CSV, Excel, MATLAB, and Python scripts, with a maximum storage capacity of 10 GB. These simulations underscore the project's commitment to detailed computational analyses and model refinements, crucial for advancing technological capabilities and performance assessments.

Furthermore, longitudinal data pertaining to stakeholder analyses, community strategies, and industrial impact assessments are structured in Excel, document, and PDF formats. These datasets, managed by various project partners, exhibit varying data rates from slow kilobytes (KB) per month for standards compilation to moderate volumes for periodic stakeholder engagement reports. This structured approach ensures that the project maintains comprehensive data integrity, accessibility, and relevance across its diverse operational domains, facilitating informed decision-making and strategic advancements throughout its lifecycle.

2.4 Data Storage and Security Protocols

In this section the emphasis is on data storage and security protocols across various datasets contributed by project partners. It includes information on storage locations, backup procedures, and measures to ensure data security and privacy, adhering to relevant regulations and best practices.

Table 5: Data Storage and Security Protocols.

Name of the dataset	Partner name	Data storage location	Data backup	Data security and privacy
Battery ageing lab data	AVL	on-premises storage	enterprise backup in place	sensitive information: primarily IP
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	on-premises storage	enterprise backup in place	
Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	SharePoint and DBL drive	Cloud backup	Data will be anonymized and GDPR compliant, informed consent collected. No sensitive information collected
Workstations analysis	DBL, TMCOMAS, AVL, FLENDER	SharePoint and DBL drive	Cloud backup	
User profiles	DBL, TMCOMAS, AVL, FLENDER	SharePoint and DBL drive	Cloud backup	
Ergonomics data	DBL, TMCOMAS, AVL, FLENDER	SharePoint and DBL drive	Cloud backup	
Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER	SharePoint and DBL drive	Cloud backup	
Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER	SharePoint and DBL drive	Cloud backup	
Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER	SharePoint and DBL drive	Cloud backup	
Microsoft Office	EITM Internal access	Local organization SharePoint		Data stored securely by EITM IT, GDPR and EU Regulation compliant.
Microsoft Form	ALL	Local organization SharePoint and Project database		Data securely managed by EITM IT, GDPR and EU Regulation compliant.
Standards	All	R3-MYDAS SharePoint	EWF's internal SharePoint	NA

Name of the dataset	Partner name	Data storage location	Data backup	Data security and privacy
Communication and dissemination material	All	R3-MYDAS SharePoint	EWF's internal SharePoint	
Stakeholder Analysis and community building strategy	All	R3-MYDAS SharePoint	EWF's internal SharePoint	
Gear tester data	FLE FI	Teams/web drive	company policy / 24h	Company GDPR policy / no sensitive data collected
Journal bearing tester data	FLE FI	Teams/web drive	company policy / 24h	
Flange connection tester data	FLE FI	Teams/web drive	company policy / 24h	
FEA simulations	FLE FI	Teams/web drive	company policy / 24h	
Time-domain simulations	FLE FI	Teams/web drive	company policy / 24h	
Test pictures	FLE FI	Teams/web drive	company policy / 24h	
Responses to user acceptance questionnaires	HUA	Locally	Netcompany SharePoint and HUA Google Drive	Out-of-consortium contacts responding to questionnaires will be anonymized
Residual stresses for induction hardened gear	Ikerlan	Locally in Ikerlan's servers	Server content is backed up every night	Personal data will not be shared
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	Fog and/or Cloud.	Data backup procedures will be determined together with the pilot leader.	No personal information will be used.
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML			
SICWELL Dataset	ITML			
Minutes of the meetings	All	Remotely	Weekly backup is sufficient	Data is just available witing project contact
Gearbox simulation models	LUT, Flender	Remotely	Weekly backup is sufficient	Data is private
Gearbox models	LUT, Flender	Remotely	Weekly backup is sufficient	
Measurement data	LUT, Flender	Remotely	Weekly backup is sufficient	

Name of the dataset	Partner name	Data storage location	Data backup	Data security and privacy
Experiment protocols	LUT, Flender	Remotely	Weekly backup is sufficient	
Scientific papers in production	LUT, Flender	Remotely	Weekly backup is sufficient	
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	Locally	Data will be saved on LUT local drives	We do not expect to collect or store sensitive data. Respondents' details will be withheld, anonymized or pseudonymized if requested
Real crankshaft dimensions (geometric data)	TMCOMAS	TMCOMAS server and TEAMS (if needed)	N/A	nothing is necessary
Robot program for scanning the part	AIMEN/TMCOMAS	locally at AIMEN/TMCOMAS server and TEAMS (if needed)	N/A	nothing is necessary
cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS	locally at AIMEN/TMCOMAS server. and TEAMS (if needed)	N/A	nothing is necessary
laser cladding parameters	AIMEN/TMCOMAS	locally at AIMEN/TMCOMAS server and TEAMS (if needed)	N/A	confidential
Robot program to weld the part	AIMEN/ZIKNES	locally at AIMEN/ZIKNES server and TEAMS (if needed)	N/A	nothing is necessary
welding quality monitoring data	AIMEN/TMCOMAS	locally at AIMEN/TMCOMAS server and TEAMS (if needed)	N/A	confidential
Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES	locally at AIMEN/TMCOMAS server and TEAMS (if needed)	N/A	nothing is necessary
final validation of the MYDAS solution	TMCOMAS	locally at TMCOMAS server and TEAMS (if needed)	N/A	nothing is necessary

AVL manages the battery ageing lab data on-premises, with enterprise-level backups in place, primarily safeguarding sensitive intellectual property (IP). Similarly, fleet data from OEMs, Tier-1 suppliers, and fleet operators is stored on-premises with robust backup mechanisms, ensuring data integrity and availability for operational analyses and optimizations.

For collaborative tasks and analyses involving DBL, TMCOMAS, AVL, and FLENDER, datasets such as tasks descriptions, workstations analysis, user profiles, and ergonomics data are stored on SharePoint and DBL drives, with cloud backups

implemented. These datasets adhere to GDPR compliance, incorporating anonymization protocols and informed consent practices, underscoring the project's commitment to data privacy and regulatory standards.

Neurophysiological data including eye tracking, EEG, and galvanic measurements, managed by the same partners, are similarly stored on SharePoint and DBL drives with cloud backups, emphasizing the secure handling of sensitive physiological information crucial for user-centric design improvements and operational insights.

Operational data from Microsoft Office applications and user feedback via Microsoft Forms are managed securely within local SharePoint databases, ensuring compliance with GDPR and EU regulations. These datasets are overseen by EITM IT policies, prioritizing data security and privacy while facilitating efficient project management and stakeholder engagements.

Technical testing and simulations conducted by FLE-FI, such as gear tester data, journal bearing tester data, FEA simulations, and others, are stored on Teams and web drives with stringent company policies for daily backups. These protocols adhere to GDPR guidelines, ensuring that no sensitive data is collected or stored unnecessarily, supporting both operational continuity and regulatory compliance.

Across all datasets, whether hosted locally or in the cloud, partners implement robust backup strategies and stringent security measures to safeguard data integrity and confidentiality. This structured approach ensures that the project maintains high standards of data governance, facilitating collaborative research and development efforts while adhering to legal and ethical guidelines throughout its duration.

2.5 Data Access and Sharing Policy

This section details the policies for data access and sharing among project partners. It outlines who will have access to the data, under what conditions, and any formal agreements or contracts governing data sharing.

Table 6: Data Access and Sharing Policy.

Name of the dataset	Partner name	Data access and sharing	Data sharing agreement
Battery ageing lab data	AVL	no sharing of raw data	Sharing derived models, not raw data.
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators		
Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	Project partners involved in activities of WP2, WP3, WP4, WP6 and WP8	Not at the moment
Workstations analysis	DBL, TMCOMAS, AVL, FLENDER		
Users profiles	DBL, TMCOMAS, AVL, FLENDER		
Ergonomics data	DBL, TMCOMAS, AVL, FLENDER		

Name of the dataset	Partner name	Data access and sharing	Data sharing agreement
Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER		
Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER		
Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER		
Microsoft Office	EITM Internal access	Only EITM	No
Microsoft Form	ALL	Data will be shared with partners	No
Standards	All	All	No
Communication and dissemination material	All		
Stakeholder Analysis and community building strategy	All		
Gear tester data	FLE FI	Shared by team's owner in FLE FI / R3Mydas partner teams / by email to other partners	Consortium agreement
Journal bearing tester data	FLE FI		
Flange connection tester data	FLE FI		
FEA simulations	FLE FI		
Time-domain simulations	FLE FI		
Test pictures	FLE FI		
Responses to user acceptance questionnaires	HUA	The raw data will be open	No
Residual stresses for induction hardened gear	Ikerlan	Ikerlan team members	No
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	Any proprietary data will be anonymised. Access will be on-premises only.	For any proprietary data, MoU will be in place before the work starts.
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML		
SICWELL Dataset	ITML		
Minutes of the meetings	All	All project participants	No
Gearbox simulation models	LUT, Flender	Selected people	Consortium agreement
Gearbox models	LUT, Flender		
Measurement data	LUT, Flender		
Experiment protocols	LUT, Flender		
Scientific papers in production	LUT, Flender		
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	Raw data will be accessible only to LUT Business School team and agreed upon project partners.	under the terms of the contract

Name of the dataset	Partner name	Data access and sharing	Data sharing agreement
Real crankshaft dimensions (geometric data)	TMCOMAS	all WP2 participants	NO
Robot program for scanning the part	AIMEN/TMCOMAS		
cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS		
laser cladding parameters	AIMEN/TMCOMAS		
Robot program to weld the part	AIMEN/ZIKNES		
welding quality monitoring data	AIMEN/TMCOMAS		
Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES		
final validation of the MYDAS solution	TMCOMAS		

AVL, managing the battery ageing lab data, adopts a policy of not sharing raw data but instead shares derived models with other partners. This approach ensures that sensitive information regarding cell ageing tests remains confidential while allowing for collaborative insights derived from processed data.

For fleet data from OEMs, Tier-1 suppliers, and fleet operators, specific details on data access and sharing policies are not specified in the provided array. However, it can be inferred that stringent protocols are likely in place to manage access and distribution due to the sensitive nature of operational vehicle data.

In collaborative efforts involving DBL, TMCOMAS, AVL, and FLENDER across various datasets such as tasks descriptions, workstations analysis, and user profiles, access is primarily restricted to project partners involved in specific work packages (WP2, WP3, WP4, WP6, and WP8). This limited access ensures that data confidentiality is maintained within the consortium, promoting focused collaboration without widespread dissemination.

Similarly, neurophysiological data including eye tracking, EEG, and galvanic measurements are restricted to the consortium members of DBL, TMCOMAS, AVL, and FLENDER, with specific sharing agreements not detailed in the provided data. This targeted approach to data access supports the ethical handling of sensitive physiological information for research and development purposes.

Microsoft Office data, accessible only to EITM internally, follows a strict policy of non-sharing to maintain confidentiality and data integrity. Conversely, data from Microsoft Forms is shared among all project partners, highlighting transparency in user feedback dissemination while maintaining privacy and compliance with data protection regulations.

For technical data such as standards, gear tester data, and FEA simulations managed by FLE-FI, sharing practices vary from team owner dissemination within FLE-FI to consortium-wide agreements for certain datasets. These protocols ensure that technical insights are shared appropriately while safeguarding proprietary information and adhering to consortium agreements.

Across all datasets, whether involving Ikerlan's residual stresses data or ITML's battery aging datasets, access and sharing policies are tailored to each dataset's sensitivity and the project's collaborative needs. These policies often include anonymization of proprietary data, on-premises access restrictions, and the establishment of Memoranda of Understanding (MoU) or contractual terms to govern data sharing agreements effectively.

The analysis of data sharing agreements among project partners reveals significant diversity in approaches and priorities regarding data access and dissemination. Several partners, such as AVL and ITML, maintain stringent controls over data sharing, reflected in their explicit "No" responses to formal agreements. This approach underscores their commitment to protecting intellectual property and sensitive information, crucial for maintaining competitive advantage and complying with regulatory standards. Conversely, responses indicating "Consortium agreement" for certain datasets, such as those managed by FLE-FI, highlight collaborative efforts among partners to establish clear guidelines and protocols for accessing and sharing data. These agreements likely facilitate smoother coordination, ensuring that all consortium members adhere to predefined rules governing data usage, confidentiality, and ethical considerations.

Moreover, the prevalence of responses indicating "No" to data sharing agreements alongside assurances of transparency or anonymization practices underscores a balancing act between collaboration and safeguarding data integrity. For instance, partners like HUA and TMCOMAS indicate a willingness to share data openly or with restricted access, contingent upon transparency measures and compliance with data protection regulations. This approach promotes trust and accountability within the consortium, fostering a conducive environment for collaborative research while mitigating risks associated with unauthorized data access or misuse. Overall, the observed outcomes highlight the intricate interplay between data governance, collaboration dynamics, and regulatory compliance in complex research initiatives like the R3-Mydas project, underscoring the importance of tailored data management strategies to achieve project goals effectively

In summary, the project's approach to data access and sharing ensures a balance between collaborative innovation and data security, facilitating meaningful research and development outcomes while adhering to legal, ethical, and contractual obligations within the consortium.

2.6 Data Quality Control Measures

This section focuses on the measures implemented to ensure the quality of the data collected and used in the project. It covers data documentation and metadata management, version control, and any processes for data cleaning and validation.

Table 7: Data Quality Control Measures.

Name of the dataset	Partner name	Data quality control	Data documentation and metadata	Data versioning
Battery ageing lab data	AVL	It is part of CRISP-DM process	Data is annotated with meta-data	is in place
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators			
Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	N/A	NA	NA
Workstations analysis	DBL, TMCOMAS, AVL, FLENDER			
Users profiles	DBL, TMCOMAS, AVL, FLENDER			
Ergonomics data	DBL, TMCOMAS, AVL, FLENDER			
Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER			
Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER			
Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER			
Microsoft Office	EITM Internal access	Deliverables peer-review processes	In internal database of EITM each project has specific folders. In the Project database that managed by the coordinator, each WP has separate folder that managed by the corresponding WP leader.	Each produced version has specific characteristics.

Name of the dataset	Partner name	Data quality control	Data documentation and metadata	Data versioning
Microsoft Form	ALL	With partners and through collaborative meetings and workshop	The survey document is stored in the internal account of EITM. The data will be shared with project partners as well.	3 version of the survey are stored as follow:
Standards	All	No	NA	Multiple data versions will be managed. Section 5.5.4.3.
Communication and dissemination material	All		NA	Multiple data versions will be managed. Section 5.5.4.3.
Stakeholder Analysis and community building strategy	All		NA	Multiple data versions will be managed. Section 5.5.4.3.
Gear tester data	FLE FI	Review of test data by engineer	Project folders / company policy	Working documents in Teams / revisioning in Teamcenter
Journal bearing tester data	FLE FI			
Flange connection tester data	FLE FI			
FEA simulations	FLE FI			
Time-domain simulations	FLE FI			
Test pictures	FLE FI			
Responses to user acceptance questionnaires	HUA	Validation of answers by HUA	any project-wide conventions	2 versions
Residual stresses for induction hardened gear	Ikerlan	No cleaning or validation	N/A	A single version is expected
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	Data management pipelines	Data documentation will follow the company's documentation policy in accordance with ISO9001.	ITML will conduct data version control, if necessary.
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML			
SICWELL Dataset	ITML			
Minutes of the meetings	All	Data will be checked during collection	File naming and folder structure agreed during data production.	Manual versioning unless data platform handles versioning

Name of the dataset	Partner name	Data quality control	Data documentation and metadata	Data versioning
Gearbox simulation models	LUT, Flender	Data will be checked during collection		Manual and automatic using git
Gearbox models	LUT, Flender			Manual and automatic using git
Measurement data	LUT, Flender			Manual and automatic using git
Experiment protocols	LUT, Flender			Manual and automatic using git
Scientific papers in production	LUT, Flender			Automatic
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	ethical protocols	To be determined	To be determined
Real crankshaft dimensions (geometric data)	TMCOMAS	N/A	N/A	add rev.XX in the name of the file
Robot program for scanning the part	AIMEN/TMCOMAS			
cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS			
laser cladding parameters	AIMEN/TMCOMAS			
Robot program to weld the part	AIMEN/ZIKNES			
welding quality monitoring data	AIMEN/TMCOMAS			
Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES			
final validation of the MYDAS solution	TMCOMAS			

The data quality control measures across partners in the R3-Mydas project reveal a structured approach to ensuring data reliability and integrity. AVL, for instance, adheres to the CRISP-DM process for battery ageing lab data, emphasizing systematic data cleaning and validation practices. Similarly, ITML implements rigorous data management pipelines for datasets like the lithium-ion battery aging profiles, aligning with ISO9001 documentation standards to maintain high-quality data. Partners such as FLE-FI employ engineer reviews and company policy-driven revisioning processes for datasets like gear tester data, ensuring that data undergoes thorough scrutiny and documentation throughout its lifecycle.

Documentation and metadata practices also play a pivotal role, exemplified by EITM's internal database and project-specific folders for Microsoft Office documents. This structured approach ensures that each data version is appropriately documented, facilitating traceability and accessibility for project stakeholders. Moreover, the use of manual and automated versioning systems, as seen with LUT and Flender for gearbox simulation models and scientific papers, underscores the commitment to maintaining data coherence and tracking changes effectively. These measures collectively enhance data reliability, support collaborative research efforts, and align with project objectives aimed at leveraging high-quality data for meaningful insights and outcomes.

2.7 Data Retention and Destruction Policies

This section outlines the policies for data retention and destruction. It describes how long data will be retained, under what conditions it will be destroyed, and the procedures for ensuring that data is disposed of securely and in compliance with relevant regulations.

Table 8: Data Retention and Destruction Policies.

Name of the dataset	Partner name	Data retention and destruction
Battery ageing lab data	AVL	Data will be retained in our own premises
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	
Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	Anonymized research data will be retained with the adequate security standards. The informed consents will be destroyed at the end of the project. All the other data will be delated upon request of the interested person.
Workstations analysis	DBL, TMCOMAS, AVL, FLENDER	
Users profiles	DBL, TMCOMAS, AVL, FLENDER	
Ergonomics data	DBL, TMCOMAS, AVL, FLENDER	
Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER	
Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER	
Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER	
Microsoft Office	EITM Internal access	The data will be stored in internal sources of EITM even
Microsoft Form	ALL	The data will be stored in internal sources of EITM
Standards	All	If new information replaces the older one, both will be maintained for one year
Communication and dissemination material	All	The information will be available at least 5 years
Stakeholder Analysis and community building strategy	All	The information will be available at least 5 years

Name of the dataset	Partner name	Data retention and destruction
Gear tester data	FLE FI	Retained as long as teams' channel is retained, final reports 25 years Teamcenter
Journal bearing tester data	FLE FI	
Flange connection tester data	FLE FI	
FEA simulations	FLE FI	
Time-domain simulations	FLE FI	
Test pictures	FLE FI	
Responses to user acceptance questionnaires	HUA	The data can be retained forever
Residual stresses for induction hardened gear	Ikerlan	Data will be retained only during project
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	ITML will not retain and/or store any proprietary data in its premises. Any publicly datasets will be retained maximum 2 years after the project ends.
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML	
SICWELL Dataset	ITML	
Minutes of the meetings	All	It will be required for the duration of the project
Gearbox simulation models	LUT, Flender	Until the end of the project, then moved to Flender
Gearbox models	LUT, Flender	Until the end of the project
Measurement data	LUT, Flender	Until the end of the project then moved to Flender
Experiment protocols	LUT, Flender	Indefinitely
Scientific papers in production	LUT, Flender	Indefinitely
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	To be determined
Real crankshaft dimensions (geometric data)	TMCOMAS	No need to destroy it
Robot program for scanning the part	AIMEN/TMCOMAS	
cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS	
laser cladding parameters	AIMEN/TMCOMAS	
Robot program to weld the part	AIMEN/ZIKNES	
welding quality monitoring data	AIMEN/TMCOMAS	

Name of the dataset	Partner name	Data retention and destruction
Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES	
final validation of the MYDAS solution	TMCOMAS	

The data retention and destruction policies across partners in the R3-Mydas project highlight varied approaches tailored to the nature of the datasets and compliance requirements. AVL, for instance, retains battery ageing lab data on their premises, ensuring control and security over sensitive information. Similarly, DBL, TMCOMAS, AVL, and FLENDER anonymize research data like tasks descriptions and neurophysiological data, with strict adherence to informed consent protocols and provisions for data deletion upon request, showcasing a commitment to data privacy and ethical standards.

EITM and all partners maintain internal storage for Microsoft Office documents and forms respectively, ensuring data accessibility within controlled environments. Standards and communication materials are retained for specific durations, such as one year for replaced standards and a minimum of five years for dissemination strategies, underscoring the importance of historical data access in strategic planning. Meanwhile, FLE-FI implements long-term retention strategies, such as retaining gear tester data for 25 years in Teamcenter, emphasizing archival practices for comprehensive data management. Conversely, ITML adopts a conservative approach, limiting retention of proprietary data to a maximum of two years post-project for publicly accessible datasets, aligning with privacy regulations and data stewardship principles.

These policies collectively reflect a balance between maintaining data integrity and adhering to legal and ethical considerations throughout the project lifecycle. Each partner's approach ensures that data is preserved securely, accessible as needed, and appropriately disposed of or archived in accordance with project-specific requirements and regulatory frameworks.

2.8 Data Analysis and Reporting

This section provides an overview of the approaches for analysing data and generating reports. It includes the methodologies for data analysis, the types of reports that will be produced, and the responsibilities for carrying out these tasks.

Table 9: Data Analysis and Reporting.

Name of the dataset	Partner name	Data analysis and reporting
Battery ageing lab data	AVL	Data will be used to develop or train models
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	
Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	DBL. Reported in D2.1, 3.1, 4.1, 5.4, 5.6

Name of the dataset	Partner name	Data analysis and reporting
Workstations analysis	DBL, TMCOMAS, AVL, FLENDER	
Users profiles	DBL, TMCOMAS, AVL, FLENDER	
Ergonomics data	DBL, TMCOMAS, AVL, FLENDER	
Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER	
Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER	
Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER	
Microsoft Office	EITM Internal access	The data will be used for preparing the project reports in all deliverables derived from the WP8 led by EITM
Microsoft Form	ALL	The data analysed through excel instruments (Pivot) and will be used for the D8.1 and D8.2
Standards	All	The data will be reported in the recommendation document to ISO/CEN. Responsible: EWF
Communication and dissemination material	All	Some data will be reported in the dissemination and communication plan. Responsible: EWF
Stakeholder Analysis and community building strategy	All	All the data will be available in the final stakeholder analysis and community building strategy. Responsible: EWF
Gear tester data	FLE FI	Data analysed by internal company tools
Journal bearing tester data	FLE FI	
Flange connection tester data	FLE FI	
FEA simulations	FLE FI	
Time-domain simulations	FLE FI	
Test pictures	FLE FI	
Responses to user acceptance questionnaires	HUA	The Behavioural Intention to Use (BIU). Responsible: HUA Output: D5.6
Residual stresses for induction hardened gear	Ikerlan	A document will be presented based on the results (responsible: Mikel Escalero)
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	Machine learning methods will be generated as part of D3.2 and D3.3. ITML's team will be responsible.
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML	
SICWELL Dataset	ITML	
Minutes of the meetings	All	Manually

Name of the dataset	Partner name	Data analysis and reporting
Gearbox simulation models	LUT, Flender	Manually
Gearbox models	LUT, Flender	Manually
Measurement data	LUT, Flender	Manually
Experiment protocols	LUT, Flender	Manually
Scientific papers in production	LUT, Flender	Manually
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	Responsible: LUT Business School. If a survey is conducted by an external agency they will initial report the data according to agreement.
Real crankshaft dimensions (geometric data)	TMCOMAS	N/A
Robot program for scanning the part	AIMEN/TMCOMAS	
cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS	
laser cladding parameters	AIMEN/TMCOMAS	
Robot program to weld the part	AIMEN/ZIKNES	
welding quality monitoring data	AIMEN/TMCOMAS	
Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES	
final validation of the MYDAS solution	TMCOMAS	

In the context of data analysis and reporting for the MYDAS project, partners employ diverse methodologies and tools tailored to their specific datasets and project goals. AVL focuses on using battery ageing lab data to develop and train models, highlighting a practical application of their data in advancing research and development. Similarly, DBL, TMCOMAS, AVL, and FLENDER integrate tasks descriptions into project deliverables such as reports for various work packages (WP), ensuring comprehensive documentation across multiple aspects of the project's scope.

EITM manages Microsoft Office data internally, primarily for preparing project reports derived from WP8, demonstrating their centralized approach to data utilization within the organization. All partners leverage Microsoft Form data through Excel instruments like Pivot tables for creating deliverables D8.1 and D8.2, underscoring their analytical approach to harnessing survey insights effectively. Standards-related data from all partners is consolidated for reporting in recommendation documents to ISO/CEN, showcasing collaborative efforts in influencing industry standards.

FLE-FI employs internal tools for analysing gear tester data, emphasizing proprietary methods tailored to their company's specific needs. HUA focuses on analysing

responses from user acceptance questionnaires, particularly in measuring Behavioural Intention to Use (BIU), crucial for understanding user perceptions and project outcomes. ITML utilizes machine learning methods to analyse lithium-ion battery aging datasets for generating deliverables D3.2 and D3.3, highlighting their expertise in advanced data analytics within the EV sector.

Overall, the R3-Mydas project integrates a wide array of data analysis and reporting techniques across its consortium, ensuring that each partner's contributions are synthesized effectively to meet project objectives and deliver actionable insights in various domains from engineering simulations to stakeholder engagement strategies.

3 Advanced Technology and Machine Learning Integration

3.1 Integration of Advanced Technologies

This section discusses the integration of advanced technologies such as blockchain, digital product passports, and international data spaces (IDS) within the project's data management framework. It explains how these technologies will be used to enhance data integrity, traceability, and interoperability.

Table 10: Integration of Advanced Technologies.

Name of the dataset	Partner name	Blockchain technology	Digital product passport	International data spaces (IDS)
Battery ageing lab data	AVL	events & data relevant for auditing, if at all	aggregated data might be relevant for battery passport	infrastructure not in place yet
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	events & data relevant for auditing, if at all	aggregated data might be relevant for battery passport	infrastructure not in place yet; EDC connector

In the realm of integrating advanced technologies within the R3-Mydas project, partners like AVL and NCI are exploring several innovative avenues. AVL's approach with battery ageing lab data involves considering blockchain technology for recording events and relevant data to enhance auditability, though the implementation infrastructure is not yet established. Additionally, they see potential in using aggregated data for digital product passports, which could streamline the traceability and lifecycle management of battery components.

Similarly, OEMs, Tier-1 suppliers, and fleet operators are focusing on leveraging blockchain for auditing purposes concerning fleet data. They also foresee the possibility of utilizing aggregated data for creating digital product passports, although the infrastructure necessary for these advancements, including EDC (Eclipse Dataspace Component) connectors, is still under development. This proactive stance highlights their commitment to integrating advanced technologies like blockchain and digital product passports to improve data transparency, auditability, and lifecycle management within the automotive sector.

Simultaneously, Task 6.3 Marketplace, led by NCI and involving partners CSEM, SPIN, LUT, ITML, DBL, and HUA, aims to implement the R3-Mydas marketplace over a duration from Month 6 to Month 36. This marketplace comprises two essential components: firstly, a platform for remanufacturing resources such as tools, materials, production line services, and consulting, integrated with the existing MARKET4.0 marketplace. Secondly, a marketplace for remanufactured products incorporating blockchain

technology for Digital Product Passports. This approach allows stakeholders to document and review the manufacturing history of products securely, with enterprise-sensitive information managed outside the blockchain and accessed through IDS-compatible data spaces.

As the R3-Mydas project progresses, these initiatives underscore the collaborative efforts among partners to harness the potential of advanced technologies such as blockchain and digital product passports, aiming to enhance data integrity, audit trails, and overall efficiency in managing and utilizing complex datasets related to battery ageing and fleet operations.

3.2 Machine Learning Model Integration

This section details the role of machine learning models in the project. It covers the names and descriptions of the models, their development and training processes, deployment and inference strategies, and maintenance procedures to ensure continuous improvement and reliability.

Table 11: Machine Learning Model Integration.

Name of the dataset	Partner name	Machine learning model name	Machine learning description	Model development and training	Model deployment and inference	Model maintenance
Battery ageing lab data	AVL	•Physical, electro-chemical laws •Semi-physical / empirical model •Statistical / machine learning	•Physical, electro-chemical laws •Semi-physical / empirical model • Statistical / machine learning	Measurement data will be read-only, no versioning planned. Script packages will be version controlled via GitHub.	OTA deployment to on-board devices; model re-training in the cloud	model cards; data sheets; CRISP-DM process
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators					
Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	• Convolutional Neural Network (CNN) • Linear Regression Long Short-Term Memory (LSTM) • Recurrent neural networks (RNNs) • Feedforward Neural Network (FNN)	Training will use public data; validation will use proprietary data. ITML's dedicated team will handle this.	Git and MLflow are possible version control mechanisms to be used for tracking changes in datasets.	Deployment will be performed using docker images.	Model maintenance includes retraining with new data. We'll use version control for models, code, and artifacts to ensure traceability and reproducibility. .
EVBattery: Large-scale EV dataset for battery health and capacity estimation.	ITML					

Name of the dataset	Partner name	Machine learning model name	Machine learning description	Model development and training	Model deployment and inference	Model maintenance
SICWELL Dataset	ITML					

Several partners are deploying diverse approaches tailored to their specific datasets and objectives. AVL, focusing on battery ageing lab data, employs physical and electro-chemical laws as the foundation for their models. These models are essential for interpreting complex datasets related to battery performance. The development and training of these models emphasize adherence to the CRISP-DM process, ensuring robustness and reliability. AVL plans to manage script packages through GitHub for version control, while deploying models via OTA (Over-The-Air) updates to onboard devices, with cloud-based re-training capabilities to enhance model accuracy over time.

For fleet data from vehicles, OEMs, Tier-1 suppliers, and fleet operators utilize statistical and machine learning models to extract valuable insights. These models leverage empirical data to predict and optimize fleet performance metrics. While specific details on model development, training, deployment, and maintenance are yet to be fully outlined, the approach underscores their commitment to enhancing operational efficiency and performance monitoring within the automotive sector.

Meanwhile, ITML focuses on advanced neural network models for datasets like lithium-ion battery aging profiles and large-scale EV datasets. For instance, they employ Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to analyse complex data patterns and predict battery health and capacity. ITML emphasizes rigorous model training with public and proprietary data, employing Git and MLflow for version control to ensure transparency and reproducibility. Deployment strategies involve using docker images for scalability, with ongoing model maintenance protocols that include regular retraining with updated data to adapt to evolving conditions and improve predictive accuracy.

Furthermore, CSEM, contributing within WP6 on quality control predictive models, implements a Machine Learning pipeline focusing on constrained Reinforcement Learning (RL) algorithms. This approach involves logging parameters and performance metrics of processes to optimize quality control measures effectively.

Across these initiatives, partners in the R3-Mydas project are leveraging state-of-the-art machine learning techniques to optimize operations, enhance predictive capabilities, and streamline lifecycle management in their respective domains. These efforts underscore the project's commitment to integrating cutting-edge technologies to drive innovation and sustainability in remanufacturing and automotive sectors.

4 FAIR Data

This Data Management Plan (DMP) outlines the procedures for managing data in alignment with the FAIR principles[2] and H2020 guidelines[3]. The acronym FAIR—standing for Findable, Accessible, Interoperable, and Reusable—highlights the essential characteristics that research data must possess to facilitate maximum knowledge circulation and return on investment. In accordance with these principles, R3-Mydas will ensure that data management practices support the discoverability, accessibility, interoperability, and reusability of research data. While some data will remain protected and not publicly accessible, the project will strive to maximize access and reuse of the generated research data.

4.1 Making Data Findable

To ensure data is findable, R3-Mydas will evaluate and utilize standard services like B2SHARE[4] or similar platforms to guarantee the creation of appropriate metadata. Whenever possible, metadata will be automatically extracted following established community standards. For datasets generated by use cases, field experts may choose to provide manually curated metadata where it is deemed more suitable.

All published data will be accompanied by comprehensive metadata features, including but not limited to the project name and Grant Agreement (GA) number. The metadata will also include:

- Abstract/description
- Access and licensing information
- Associated project and community
- Associated publications and reports
- Bibliographic information
- Digital Object Identifiers (DOIs)
- Grant information
- Keywords
- Language
- Version numbers

[2] *Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016).* https://doi.org/10.1038/sdata.2016.18

[3] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

[4] https://b2share.eudat.eu/

4.2 Making Data Openly Accessible

For peer-reviewed scientific publications resulting from the R3-Mydas project, open access publication will be pursued in compliance with the R3-Mydas Grant Agreement (GA). This entails depositing a machine-readable electronic copy of either the published version or the final peer-reviewed manuscript accepted for publication in a repository for scientific publications. Additionally, the accompanying research data required to validate the results presented must be deposited as soon as possible.

Open access to the deposited publication via the repository must be ensured by the time of publication if the publisher provides a free electronic version. Otherwise, access must be granted within six months of publication (or twelve months for publications in the social sciences and humanities). The bibliographic metadata identifying the deposited publication must also be openly accessible in a standard format. This metadata must include the terms "European Union (EU)" and "Horizon Europe", the action's name, acronym, and grant number, the publication date, any applicable embargo period, and a persistent identifier.

Concerning research data, efforts will be made to make them available as much as possible, in alignment with the R3-Mydas GA. However, for use case-specific data, it is currently unclear which data will be generated, making it difficult to specify which data will be made publicly available. For (sensitive) personal data, especially in the medical use case, any real patient data utilized for training purposes will be anonymized before being made publicly accessible. On the other hand, data that is sensitive to the business interests of the use case providers will be protected. Stakeholder-related data, predominantly containing personal information, will not be made publicly available. Nonetheless, anonymized results from workshops and other stakeholder engagement activities will be published in related deliverables.

The datasets produced within R3-Mydas will be made openly available through platforms and repositories such as Zenodo[5] and OpenAIRE[6], accompanied by proper documentation and metadata describing the data features. Zenodo, an open-access repository developed under the European OpenAIRE program and operated by CERN, supports the sharing, curation, and publication of research data and software. OpenAIRE, commissioned by the European Commission, provides a comprehensive repository for EC-funded research to support the Open Data policy. Zenodo also facilitates data versioning by automatically assigning a Digital Object Identifier (DOI) to every publicly available upload, ensuring that each upload is uniquely citable.

[5] https://zenodo.org/
[6] https://www.openaire.eu/

4.3 Making Data Interoperable

R3-Mydas will utilize data interoperability standards from the industry 4.0[7] and manufacturing[8] sectors to ensure that the data produced by the R3-Mydas use cases meets the interoperability criteria specific to these domains.

While it is understood that not all datasets collected and generated throughout the project will be public, certain confidential datasets will be safeguarded and will not be accessible or interoperable. Nevertheless, datasets that are designated for public availability will be deposited in open access repositories like Zenodo, where they can achieve interoperability. Zenodo utilizes the JSON (JavaScript Object Notation) schema for managing metadata, thereby ensuring that these public datasets are well-structured and easily integrated with other systems.

4.4 Making Data Reusable

Data generated by R3-Mydas will be distributed under Creative Commons Attribution Licence (CC-BY) licenses, enhancing reusability by ensuring the data is freely reusable while maintaining traceability and credit to the original data providers. If data are created or used with proprietary software (e.g., from use case partners), the documentation accompanying these datasets will recommend alternative and compatible open-source tools. Additionally, datasets shared on open data repositories will be connected to related research publications on the project's website and will be updated based on their application in subsequent research that extends R3-Mydas.

[7] https://www.iiconsortium.org/IISF/

[8] https://www.mimosa.org/

5 Allocation of Resources

Costs associated with research data management and making data open are eligible under the project grant. A first estimation is that up to 5% of the total project costs will be necessary to ensure the research data is quality-controlled, FAIR-compatible, and as open as possible. Throughout the project, consortium partners will manage and curate the datasets in their possession. Upon project completion, the consortium steering group will assign the responsibility of long-term preservation and sharing of datasets to the Principal Investigator or the project data manager.

R3-Mydas utilizes free research data repository tools. The project budget includes allocations for data management costs, as most partners have accounted for Open Access fees in their budgets, ensuring no additional costs are anticipated for this activity.

6 Ethical Data Management and Compliance Strategy

6.1 General Data Protection Regulation (GDPR) Compliance

The General Data Protection Regulation (GDPR)[9] was formally introduced in May 2018 and has since been enforceable across all Member States of the European Union, as well as in the European Economic Area (EEA) countries.

Ensuring data confidentiality is a paramount concern throughout the R3-Mydas project and its subsequent use. The developed solution within R3-Mydas is intended for ongoing utilization, necessitating strict compliance with GDPR regulations. Stakeholder data collected for the project will adhere to relevant ethical standards and local data collection requirements. It will be securely processed and managed in accordance with applicable privacy laws and data protection regulations.

6.2 General Data Protection Policy

The subsections below provide and describe the R3-Mydas Policy for General Data Protection.

6.2.1 Introduction

This General Data Protection Policy (the "**Policy**") is drafted by NCI (the "**Project Coordinator**") with regard to the EU Horizon Europe Project R3-Mydas Grant agreement ID 101138738 (the "**Project**") executed by the list of partners included therein (the "**Project Partners**") in order to:

- Comply with the policy and legal requirements of the EU General Data Protection Regulation (Regulation EU 2016/679, the "**GDPR**")14, as in effect since 25 May 2018;
- Comply with all other applicable national and EU regulations and guidelines on personal data processing;
- Comply with applicable regulations and best practices with regard to research projects within the EU HE Research Programme;
- Raise awareness and improve knowledge among the Project Coordinator, the Project Partners, as well as their employees and/or agents and/or contractors (collectively, the "**Policy Recipients**").

Because the field of data protection is a dynamic legal field of constant change, new developments, in the form of new regulations, official reports and/or guidelines, are issued by EU and national legislators, as well as competent national authorities at a

[9] https://gdpr-info.eu/

constant pace. In this context, this Policy may need to be periodically updated by the Project Coordinator, in order to remain relevant to legislative change. Accordingly, Policy Recipients will be duly informed, and will be asked to provide their renewed consent upon any such updates.

6.2.2 Definitions

For the purposes of this Policy the GDPR definitions, as set in Article 4, apply. In addition,

"**Personal data**" means any information relating to an identified or identifiable natural person that is processed by any Project Partner and Policy Recipient during execution of the Project.

"**Controller**" means the owner of the personal data (usually the creator of the data itself), unless otherwise expressly clarified in this Policy or elsewhere in Project deliverables and/or reports.

"**Processor**" means each Project Partner, unless otherwise expressly clarified in this Policy or elsewhere in Project deliverables and/or reports.

"**Consent**" of the data subject means any freely given, specific, informed, unambiguous and in writing indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

"**Supervisory authority**" means the competent Data Protection Authorities within the Project Partners' jurisdictions.

The aim of the above definitions is to particularise and complement the definitions of Article 4 of the GDPR. Policy Recipients are advised to consult both texts in order to formulate the applicable definitions each time.

6.2.3 Policy Scope

The Controller determines in advance what is the law applicable to the processing of personal data in a particular case, considering that according to EU law such determination comes from legal principles and cannot be derogated by the parties.

6.2.3.1 Establishment

Each Project Partner is established on the territory of EU Member States. In the event of any change in establishment, the respective Project Partner shall notify the Project Coordinator duly and in writing.

Unless otherwise expressly specified, each Project Partner is considered the Controller in that Member State.

6.2.3.2 Processor outside the EU

In the event of any subcontracting to an organization not established on EU territory (such as subsidiaries pertaining to the same corporate group) that processes personal data of people staying on EU territory, on behalf of a Project Partner, that organization qualifies as Processor and ensures the fulfilment of the obligations imposed by the GDPR for that specific part of processing.

6.2.4 Personal Data Processing

The subsections that follow describe the policy relating to personal data processing. For consistency and completeness, it should be noted that in the R3-Mydas project, there is no processing of personal data whatsoever.

6.2.4.1 Personal Data

Personal data means any information relating to natural persons, which is or can be identified, even indirectly, by reference to any other information including a personal identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

6.2.4.1.1 Special Categories of Data

Special categories of personal data include data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, data concerning health or data concerning a natural person's sex life or sexual orientation as well as the processing of genetic data and biometric data for the purpose of uniquely identifying an individual.

In the event of such processing, the Controller and/or Processor should follow special rules related to, such as acquiring specific informed consent from the data subjects and applying stricter safeguards.

When the Controller and/or Processor relies on the data subject's consent as a legal ground for processing special categories of data, they will meet all legal consent requirements; otherwise, the data may only be processed if and to the extent that it is based on one of the legal grounds listed in the GDPR for the processing of such data.

6.2.4.1.2 Data Anonymization

Whenever possible, including non-detrimental to Project execution purposes, the Controller and the Project Partners shall undertake efforts to keep personal data processed by them for project purposes anonymous or pseudonymous.

According to the GDPR, "anonymous information" is information which does not relate to an identified or identifiable natural person, or personal data that are rendered anonymous in a way that the data subject cannot be identified. In this context, the GDPR does not apply to the processing of such anonymous information, including for statistical or research purposes.

Similarly, "pseudonymisation" means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

6.2.4.1.3 Newsletters, Social Media and other Dissemination Material

Unless otherwise specified in the Project contract, the Controller shall be responsible for the personal data processing carried out for Project dissemination purposes. To this end, the Controller shall:

- Collect and keep all relevant personal data (including lists of contact details), or copies thereof;
- Monitor relevant communications;
- Issue to Project Partners instructions and guidelines on Project dissemination activities (including any EU or other state guidelines, whenever available);
- Inform the Project Partners of any policy or legal requirements, reviews and changes.

6.2.4.1.4 Minors

Processing of children's personal data requires a special legitimate basis. In the event of such processing the Controller shall be informed in advance and in writing by Project Partners.

6.2.4.2 Data Processing

Data processing means any operation, or set of operations, carried out with or without the help of electronic or automated means, concerning the collection, recording, organization, storage, interrogation, elaboration, modification, selection, retrieval, comparison, utilization, interconnection, blocking, communication, dissemination, erasure and destruction of data.

6.2.4.3 Principles for Legitimate Processing

The European Union data protection law set forth the following specific principles which have to be complied with for the processing to be legitimate.

Pertinence and necessity - The Controller should implement management practices to fulfil the obligation to collect only relevant and necessary data for a specified purpose.

Purpose limitation - Personal data is collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. The Controller has a clear overview of all purposes for which personal data is processed. Personal data is not processed for purposes besides the original purposes, unless the (secondary) use is compatible.

Data minimization - Personal data collected by the Controller must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are

collected and further processed; if the same purposes can be realized in a less data intensive way a preference is given to that method.

Data update - Personal data is accurate, and, where necessary, kept up to date. Every reasonable step is taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.

Data retention - Personal data is kept in a form which permits identification of data subjects for no longer than it is necessary for the purposes for which the personal data are processed. The Controller and/or the Processor concerned should have processes and policies in place to:

1. determine what the applicable (minimum and maximum) retention periods are for the personal data that is being processed;
2. ensure that relevant retention periods are monitored.

6.2.5 Data Protection Legal Roles

6.2.5.1 Controller

By determining the purposes and means of the processing of personal data, unless otherwise expressly specified in this Policy, the Controller is considered by law as the "Controller" and is the primary target of the provisions of the law.

6.2.5.1.1 Identification

The data Controller previously identifies itself as such and ensures an effective implementation of data protection measures, in order to comply with the principle that personal data are processed fairly and lawfully. The legal role of the Controller implies specific responsibilities because provisions setting conditions for lawful processing are essentially addressed to the Controller.

6.2.5.1.2 Accountability

The GDPR provides full accountability of the company/Controller regarding the compliance of its processing of personal data with the law. To ensure the effectiveness of that obligation, it prompts the Controller to follow an overall approach, achieving a genuine system of control and management of its pertinent information. So, accountability and compliance systems are elements of the framework for the protection of personal data, in the cause / effect relationship: to be compliant and able to prove it (accountability), the Controller needs to put in place a comprehensive compliance system.

6.2.5.1.3 Data Protection by Design

The Controller considers data protection issues from the outset and from the design of the Project, within the whole lifecycle of processing, in order to manage the issues in a proactive way, to reduce costs and improve efficiency.

6.2.5.1.4 Data Protection by Default

The Controller standardizes data protection principles in personal data processing, products and services. The measures adopted ensure that:

- personal data is processed for purposes not different from the original purposes,
- only data necessary for these purposes are collected, and
- data are not disclosed without human intervention.

6.2.5.2 Joint Controller

If at any time during the Project execution the Controller processes personal data in conjunction with a third party, by jointly determining the purposes and means of the processing, they both act as joint Controllers. Both joint Controllers determine the mutual responsibilities with a specific arrangement.

6.2.5.3 Processor

Unless otherwise specified expressly in this Policy, all Project Partners act as Processors during Project execution.

A processor processes personal data on behalf of the Controller – that is, the Controller delegates all, or part of the processing activities to them. In such an event the Project contract assumes the role of the relevant required written agreement as per GDPR requirements.

The Processor warrants that it shall provide sufficient guarantees to ensure compliance with the GDPR, has implemented appropriate controls to meet data protection requirements defined by the agreement, instructions and/or legal requirements and ensures the protection of the rights of data subjects.

6.2.5.3.1 Auditing

The Controller ensures the commitment of the Processor(s) to enable and contribute to any review activities, including inspections, conducted by the Controller or other (EU authorities') auditors and/or reviewers, as appropriate.

6.2.5.3.2 Security

Each Project Partner undertakes that it adopts appropriate security measures to ensure the security, integrity and confidentiality of personal information and electronic communications at an adequate level with regard to Project purposes, and at any event at no lower level than processing of similar data within its own organisation.

6.2.5.4 DPO

Whenever required, following applicable GDPR and Member State respective legal requirements, the Project Coordinator (NCI) may designate a Data Protection Officer ("DPO") for assistance in monitoring internal compliance with the GDPR.

6.2.5.4.1 Identification

Each Processor appoints a DPO in accordance with the criteria and the requirements set forth in the GDPR, as applicable to it. In such an event, it shall notify the Controller in writing accordingly.

6.2.5.4.2 Designation compulsory vs. voluntary

Each Processor documents the reasons supporting the designation of the DPO or, rather, the reasons why such designation is deemed not necessary. This documentation forms part of the data protection documentation system of that Processor.

6.2.5.4.3 Professional requirements

The DPO has sufficient authority, professional qualities and independence to ensure success in his role, according to the GDPR provisions.

6.2.5.4.4 Tasks

The organization assigns to the DPO at least the tasks listed in the GDPR.

6.2.5.4.5 Notification to Supervisory Authority

Whenever a DPO is appointed, the organization notifies the Supervisory Authority of such designation and publishes DPO's contact details.

6.2.5.5 People in charge of Processing

Individuals who process personal data under the authority of the Controllers or Processor(s) must receive specific formal instructions. Hence, the Controller gives specific instructions, relating also to the implementation of security measures and safeguards, to all its personnel in charge of processing personal data.

6.2.5.5.1 Training and Awareness

All Project Partners' employees should be well informed and aware of data protection implications and be able to carry out their obligations in their work. A data protection education and communication program should be in place and supported by a monitoring system that confirms all employees and/or contractors are appropriately trained on their data protection responsibilities.

6.2.5.5.2 Policies and Procedures

Data protection policies and procedures exist, are documented in writing, are formally approved by management, implemented, reviewed, updated and approved when there are changes to applicable laws and regulations.

All Project Partners understand, and the Controller may ask them to overview all their personal data processing, the data protection risks and the applicable rules and procedures. In such an event, they shall provide it with all requested information to the best of their ability without undue delay.

6.2.6 Notice and Consent

6.2.6.1 Notice

Each Controller and/or Processor, as appropriate, provides the information required by law to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language.

The data protection notice informs data subjects about the processing of personal data relating to them, even when the personal data is not collected from them as well as of their rights, in order to let them verify in particular the accuracy of the data and the lawfulness of the processing.

6.2.6.2 Informed Consent

Explicit consent is an additional safeguard which may be required in some research studies to comply with applicable regulations. Whenever personal data is collected from participants, their informed consent must be sought in a way that meets the minimum standards of the GDPR. This requires consent to be given by a clear affirmative act. For consent to be informed, the participant must be provided with detailed information about the envisaged data processing in an intelligible form.

In Annex 1 is a template consent form which can be customized to be used by the R3-Mydas studies. It covers the main aspects that the research participant should be informed of and consent to:

- purpose of the study and key investigators involved;
- potential risks and discomforts;
- potential benefits and compensation; confidentiality terms;
- right to withdraw and where to ask any questions.

When customizing it, the language should be as clear and concise as possible.

6.2.6.3 Withdrawal of consent

Data subject's consent can be withdrawn at any time; even though it will not affect the lawfulness of processing based on consent before its withdrawal.

6.2.7 Rights of data subjects

The individual whom the data refers to (data subject) is entitled with specific rights set forth by the law. The GDPR requires that each Controller and/or Processor, as appropriate, must facilitate the exercise of the data subject's rights, take action on the request within a specific time frame and must communicate the information requested in an intelligible and easy to access form.

6.2.7.1 Right of Access

Any individual must be able to exercise the right of access to data relating to him which are being processed.

6.2.7.2 Right to Rectification

Each Controller and/or Processor, as appropriate, should have a procedure in place for data subjects to request rectification of their personal data. The procedure specifies in which cases rectification is legitimate.

If a data subject's request for rectification is legitimate, this is executed across all relevant data storage facilities, including those managed by third parties.

6.2.7.3 Right to Erasure

Each Controller and/or Processor, as appropriate, should have a procedure in place for data subjects to request erasure of their personal data. The procedure specifies in which cases erasure is legitimate.

If a data subject's request for erasure is legitimate, this is executed across all relevant data storage facilities, including those managed by third parties.

6.2.7.4 Right to Restriction of Processing

Each Controller and/or Processor, as appropriate, should have a procedure in place for data subjects to request restriction of processing of their personal data. The procedure specifies in which cases restriction is legitimate.

If a data subject's request for restriction of processing is legitimate, this is executed across all relevant data storage facilities, including those managed by third parties.

6.2.7.5 Right to Data Portability

Each Controller and/or Processor, as appropriate, determines which processes are subject to the right of data portability as well as when the requirements for such right are met.

Data subjects can request from each Controller to receive a machine-readable copy of the personal data the Controller holds about them and where possible, enable the transfer of this data to another data Controller.

Portability right can be exercised when:

- processing operations are based on data subject's consent or on contract
- personal data concerns the data subject and are the same that the latter has provided to the organization
- the right does not adversely affect rights and freedoms of others
- The processing is carried out by automated means.

Each Controller and/or Processor, as appropriate, implements appropriate measures and procedures to provide data subject, who is entitled to, with a structured, commonly used and machine-readable copy of the personal data it holds about him and where possible, to enable the transfer of this data to another data Controller indicated by data subject.

6.2.7.6 Right to Object

Where personal data are processed for scientific or historical research purposes or statistical purposes, the data subjects have the right to object on grounds relating to their particular situation (unless the processing is necessary for the performance of a task carried out for reasons of public interest). The right to object is explicitly brought to the attention of the data subject at the latest at the time of the first communication with the data subject, presented clearly and separately from any other information. Measures should be in place to assess such objections and to ensure that such processing ceases when the request is legitimate and needs to be respected.

Data subjects have the right to object, on request and free of charge, to the processing of personal data relating to them for purposes of direct marketing.

6.2.7.7 Automated Decision Making

Data subject has the right to object to any automatic decision-making (including profiling).

Each Controller and/or Processor, as appropriate, will have determined which processes entail automated decision-making (including profiling) and will have established measures to allow data subjects to object to such automated decision making and profiling. Suitable measures are in place to safeguard the data subject's rights and freedoms and legitimate interest, at least the right to obtain human intervention on the part of the Company/Controller, to express his or her point of view and to contest the decision.

6.2.7.8 Timely Response to Exercise of Rights

Each Controller and/or Processor, as appropriate, must confirm to data subjects without delay whether data relating to them are processed and communicate the data to them in an intelligible form. Each Controller and/or Processor, as appropriate, should implement internal procedures in order to be able to provide a timely response to the requests of data subject for the exercise of his rights.

Measures have to be implemented in a way that effectively allows an individual to exercise his or her right to personal data, and that enables Each Controller and/or Processor, as appropriate, to respond to such request appropriately within the required timeframes.

6.2.7.8.1 Notification to Recipients

In case of a legitimate exercise of rights to rectification, erasure or restriction of processing recipients of the personal data should be informed of the rectification, erasure of that data or of the restriction of processing.

Each Controller and/or Processor, as appropriate, should have a procedure in place for communicating any rectification or erasure of personal data or restriction of processing

to the recipients to whom the personal data has been disclosed and for disclosing these recipients to the data subject, if so requested.

6.2.8 Data Protection Documentation System

6.2.8.1 Register of Processing

Each Controller and/or Processor, as appropriate, regarding their processing activities must set up a relevant record, maintained in writing (including in electronic form) and made available easily and swiftly to the supervisory authority on request, as per applicable legal requirements within their respective Member States. The record of processing activities shall contain all the information required by GDPR.

Consequently, the Controller shall have an up-to-date overview of all personal data processing activities and shall maintain records within the Project that meet the legal requirements posed by the GDPR. By doing so, the Controller will be able to demonstrate compliance to any Supervisory Authority or other state or EU authority concerned.

For the avoidance of doubt, each Project Partner carries the same responsibility above within its own respective organisation.

6.2.8.2 Register of Data Breaches

A specific register where the breaches have to be recorded together with other information specified by the law, must be maintained by the Controller and shown to the Supervisory Authority upon request. This register is an important element of the data protection documentation system.

Project Partners need to notify immediately and in writing the Controller of any personal data breach within their respective organisations that affects execution of the Project in any way, and to cooperate with the Controller while applying relevant GDPR legal requirements.

6.2.9 Data Protection Assessment

6.2.9.1 Assessment

In the event that a Data Protection Impact Assessment ("DPIA") is carried out under the Project, the Controller shall ensure that personal data receives the appropriate level of protection in accordance with the assessed data protection risk.

The decision whether to carry out a DPIA under the Project, unless undertaken in respective Project contract, will be made by the Controller upon prior written consultation with the Project Partners.

6.2.9.1.1 Adequacy of Protection

The Controller, assisted by Project Partners, should have a process in place in order to assess for all processing the risks of varying likelihood and severity for the rights and freedoms of natural persons, taking into account the nature, scope, context and purposes of personal data processing.

6.2.9.1.2 Impact Assessment in case of High Risk (DPIA)

When the preliminary assessment highlights that processing represents high risks, a formal and documented DPIA is carried out by ascertaining possible impact on data subject.

DPIA is conducted in such a way to meet all the requirements set forth by the GDPR (art. 35) in order to confirm the quality and validity of the findings.

6.2.9.1.3 Prior Consultation to Supervisory Authority

The Controller has a process in place and roles are assigned in order to ensure that when a DPIA determines that the processing represents high risks, the competent Supervisory Authority is consulted prior to the processing.

6.2.10 Technical and Organizational Measures

The Controller and each Project Partner, as appropriate, adopts appropriate technical and organisational measures with regard to Project execution (the "Measures"), and reviews and updates them where necessary, to ensure and to be able to demonstrate that processing is in compliance with GDPR.

Each Project Partner shall notify relevant Measures to the Controller in writing. In the event of any queries or further requests by the Controller, each Project Partner undertakes to address them duly and in writing.

In the event that any Project Partner has notified the Measures to its competent Supervisory Authority, it shall inform the Controller thereof, and shall provide respective copies thereof.

6.2.11 Data Breach

According to GDPR, the Controller and/or Processor, as appropriate, has to implement adequate Measures in order to prevent personal data breaches.

In addition, the Measures should be able to minimize the adverse effects in case a security breach to personal data relating in any manner to the Project occurs anyhow.

Should a data breach occur, GDPR sets forth that the Controller and/or Processor, as appropriate, has to notify it to the Supervisory Authority providing specific information, without undue delay and in any case no later than 72 hours from the time of knowledge.

When the breach leads to significant risk of serious adverse effects on the data subject(s) or serious adverse consequences for the protection of personal data, also the latter must be informed without undue delay.

6.2.12 Data Transfers to Third Countries

No international transfers of personal data are expected to take place under the Project.

In the event that any Project Partner wishes to carry out such personal data processing in a third country, it shall notify the Controller in writing and in advance. Unless otherwise expressly specified, any international data transfers carried out by any Project Partner for any reason during Project execution take place at its own exclusive liability and responsibility; same Project Partner shall hold all other Project Partners (including the Controller) harmless from any legal or other claims arising for such personal data processing.

In case of violation of data protection principles and rules, each Project Partner (including the Controller) is responsible for damages and is subject to sanctions. Possible violations may involve civil liability and sanctions in order to ensure that any relevant damage is compensated.

The Project Partner (including the Controller) that is liable for said damages and/or sanctions shall hold all other Project Partners harmless from any claims, costs, and expenses arising from relevant GDPR infringement.

6.2.13 R3-Mydas Repository Personal Data Protection and Privacy policy

The following Personal Data Protection and Privacy Policy is uploaded onto the Project website and SharePoint:

1. Introduction. This Personal Data Protection and Privacy Policy (the "**Policy**") aims at providing details of the processing, and related methods of use, of personal data referred to users/visitors (the "**User(s)**") of this website that can be reached at the address https://r3-mydas.eu/ (the "**Website**").

This Policy refers to the EU Project [R3-Mydas, 101138738] (the "**Project**").

Web users and visitors are recommended to read this Privacy before sending any personal information and/or filling in any electronic form posted on this website.

This information is given in accordance with applicable EU data protection law, in particular the EU General Data Protection Regulation, and EU applicable Privacy law.

2. Controller. The Controller is the actual data owner per data case i.e., it is expected to be an R3-Mydas partner that has full ownership or is the creator of the dataset.

3. Scope. This Policy covers this web site only, and no other personal data processing under the Project or any other websites owned or run in any manner by the Controller or Project Partners.

4. Policy and information notice. This site has been designed with the main function of providing information on the activities of the Project. Therefore, in most cases, the collection of the user's personal data is not required.

In certain instances, such as the "newsletter" section and in order to allow the transmission of our newsletter, the interested user is required to fill out a data collection form. In these cases, the user is always free to provide his/her own data and consent to relevant processing. We recommend reading this Policy before providing the data.

In addition, should it be necessary in limited cases to collect personal information for other purposes, this will be clearly shown in the information privacy notices required by law, in order to enable transparency and user awareness. Consent forms and other documentation will be used each time, as appropriate.

The above information aims to define limits and methods of personal data processing of each service, according to which the visitor can freely express his consent and eventually allow the collection of data and its subsequent use.

5. Traffic data. The computer systems and software procedures used to operate this website acquire, during their normal operation, some personal data whose transmission is implicit in the use of Internet communication protocols.

This category of data includes: IP addresses, browser type, operating system, the domain name and website addresses from which you are logged in or out, the information on pages visited by users within the site, the time of access, time period of user's staying on a single page, the internal path analysis and other parameters regarding the user's operating system and computer environment.

This technical / IT data is collected and used only in an aggregated and not immediately identifiable manner; they could be used to ascertain responsibility in case of hypothetical crimes against the site or upon public authorities' request.

6. Cookies. No cookies are used by this repository.

7. Redirects to other websites. From this website, you can connect through special links to other websites of Project Partners within the Project, or of third parties as applicable each time. Controller hereby assumes no responsibility regarding the possible processing of personal data by third-party sites and in respect of the management of authentication credentials provided by third parties.

8. Purposes of processing and data retention. The processing of personal data is carried out mainly by using electronic procedures and media for the time strictly necessary to achieve the purposes for which the data were collected. The User, however, has the right to obtain the cancellation of his data for legitimate reasons.

9. Optional supply of personal information. The supply of personal data required from the User, unless otherwise noted, is optional, but in case of refusal it could be impossible to fulfil the request, or the related activity mentioned.

10. Place of personal data processing. Data processing related to web services of this website takes place, unless otherwise expressly stated, at Controller's establishment,

which provides for the corresponding repository management. Personal data are only handled by technical personnel of the Controller, specifically in charge of processing, or others charged with occasional maintenance operations. These persons have received specific instructions on the confidentiality of this data.

11. Scope of data flow and dissemination. The data may be used by the Controller and/or the Project Partners' employees, as persons in charge of processing, to whom appropriate operating instructions have been given, as well as by third parties who perform operating activities on behalf of them and who act as data processors, in order to fulfil contractual obligations regarding the Project. Personal data are not disseminated to unspecified recipients. Detailed information on the names of the data processors can be requested by writing to the project coordinator.

12. Data protection rights. With regard to the processing of personal data, User has the right, at any time, to obtain confirmation of whether or not the data exists and to have it communicated to him/her in an intelligible format. Users also have the right to know the content and the origin of the data, to check its accuracy or to ask that it be integrated, updated or adjusted. Finally, Users have the right to ask that the data be deleted or made anonymous or to request the blocking of data processed in violation of the law; moreover, they may oppose the processing of the data for legitimate reasons. Requests should be addressed to the project coordinator.

13. Policy updating. The possible entry into force of new laws, as well as the evolution and updating of User services or developments in the Project could make it necessary to vary the method of processing of personal data. It is therefore possible that our policy may be modified over time and therefore we invite the visitor to periodically visit this page. To this end, the policy document highlights the date of the last version.

6.2.14 R3-Mydas Day-to-Day Data Usage and Related Processes

While R3-Mydas does not handle direct personal data resulting from or processed during its research activities, it acknowledges the necessity of establishing procedural policies to ensure consensus on the use, storage, retention, opt-out procedures, and other aspects relevant to everyday project operations. The following list outlines these matters, detailing the consortium's unified approach in addressing them.

6.2.14.1 R3-Mydas List of Contacts

The R3-Mydas contact list is contained in a single XLS file that includes the names and email addresses of all consortium partners and their representatives. It also specifies the purpose of contacting each individual within their organization (e.g., administrative, technical, legal) and the email lists to which they belong. Access to this contact list is restricted to R3-Mydas consortium partners. The list is intended to facilitate organized communications within R3-Mydas. The data will be deleted after the project's conclusion

(January 2027) and the completion of the final project review and payment and will not be retained. This list is stored on the SharePoint. Individuals have the right to opt out of this list by sending an email directly to the project coordinator.

6.2.14.2 Meetings' related Material

This pertains to any document produced and utilized for project meetings, including agendas, presentations, minutes, signature lists, and other internal documents created for R3-Mydas meetings. These documents will be generated and maintained for the internal use of R3-Mydas. They will be retained for five years following the project's conclusion (until January 2027). Individuals have the right to opt out of being mentioned in these documents by sending an email directly to the project coordinator before or after the meeting.

6.2.14.3 Workshops/Conferences and Training sessions

This data pertains to the creation of workshops, including agendas, programs, participant lists, and other dissemination materials related to R3-Mydas organized events. For external publication, this material will be fully anonymized to exclude any personal information about presenters and participants in the publicly shared programs and agendas. Internally, for workshop organization within R3-Mydas, these files will be stored in the R3-Mydas SharePoint under the meetings section. The data will be retained for five years after the project's end for auditing purposes (i.e., until January 2027). Individuals have the right to opt out of being mentioned in these materials by sending an email directly to the project coordinator before or after the event.

6.2.14.4 Reporting

Reporting encompasses both internal and external (EC) documents, including progress updates and technical overviews for R3-Mydas. These files will contain documents (reports without personally identifiable information) and financial data (C forms) that may include personal data. The primary purpose of this data is financial, enabling partners to claim budget allocations for their efforts in R3-Mydas. C forms will be managed exclusively by the project coordinator and stored securely on an internal SharePoint site. This data, specific to each partner, will not be shared with anyone inside or outside of R3-Mydas, will be retained for five years post-project (for audit purposes until January 2027), and will be deleted afterward. Partners can opt out of data inclusion, but this requires submitting an updated Form C.

6.2.14.5 Deliverables, internal documents and other R3-Mydas reports

Throughout the duration of the R3-Mydas project, a substantial amount of documentation and reporting will be generated, covering project deliverables and internal documents. These files will be used to fulfil the project's contractual obligations and shared with R3-Mydas partners, the European Commission (EC), or the public, depending on the deliverable type. Author names or emails may be included in these

documents. Internal documents distributed to R3-Mydas and the EC will be used solely for reporting purposes. Publicly shared reports (public deliverables) will only mention the partner's name, excluding other personal information. All reports will be retained for five years after the project's conclusion for auditing purposes.

6.2.14.6 Source Codes

Regarding the inclusion of personal information in source code, R3-Mydas plans to avoid embedding any such information in the source code files created within the project. If any partner intends to include personal information, they must create and use a consent form, which must be signed by the data owner(s).

6.2.14.7 Usage of Cookies (in R3-Mydas Sites)

If any R3-Mydas web application requires the use of cookies, a pop-up window must inform the user and prompt them to accept or decline the conditions under which their personal information is stored. The cookie policy for the R3-Mydas website is available here. R3-Mydas will strive to minimize the use of cookies in its web development efforts.

6.2.14.8 Lists of Stakeholders and R3-Mydas Contacts

This list pertains to internal R3-Mydas records of external stakeholders, including potential technology/results adopters, significant connections with end-users, and other stakeholders. It will be used solely for R3-Mydas communication purposes and will be accessible only to R3-Mydas partners. When individuals are added to this list, they must provide consent via email. The data will be retained until the conclusion of R3-Mydas in January 2027. Any individual can opt out of being included in this list by sending an email directly to the project coordinator.

6.2.14.9 Project related Research Data

All data shared internally within R3-Mydas for research purposes must be completely anonymized by the data owner (acting as the data Controller) and must not contain any personal information, as outlined in the preceding sections.

6.2.14.10 Any other R3-Mydas-related Data

If personal information must be included in any document within R3-Mydas, the Controller (document creator) is required to inform the data owners about the inclusion of their personal details in the document, along with the purpose, retention, and storage information.

6.3 Data Governance and Ethics

This section outlines the governance frameworks and ethical considerations for managing data in the project. It includes guidelines for data integration, ethical use of data, compliance with regulations, and strategies for mitigating potential biases in machine learning models.

Table 12: Data Governance and Ethics.

Name of the dataset	Partner name	Data Integration	Data governance and ethics	Compliance and regulations	Risk Management
Battery ageing lab data	AVL	engineering teams: FMU/FMI; consume APIs via standardized protocols like HTTP	Compliance with automotive standards and horizontal regulations.	EU battery regulation; Data Act; Data Governance Act; AI Act	Revealing IP
Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	engineering teams: FMU/FMI; consume APIs via standardized protocols like HTTP	Compliance with automotive standards and horizontal regulations.	EU battery regulation; Data Act; Data Governance Act; AI Act	Revealing IP; in the B2B fleet context; not that much a privacy issue
Standards	All		EWF		no risks
Communication and dissemination material	All		EWF		no risks
Stakeholder Analysis and community building strategy	All		EWF		no risks
Gear tester data	FLE FI		Research engineer and project group		Data secured, up to date and valid
Journal bearing tester data	FLE FI		Research engineer and project group		Data secured, up to date and valid
Flange connection tester data	FLE FI		Research engineer and project group		Data secured, up to date and valid
FEA simulations	FLE FI		Research engineer and project group	Software licences	Data secured, up to date and valid
Time-domain simulations	FLE FI		Research engineer and project group	Software licences	Data secured, up to date and valid
Test pictures	FLE FI		Research engineer and project group	Company policy	Data secured, up to date and valid
Residual stresses for induction hardened gear	Ikerlan		Ikerlan		None identified

Name of the dataset	Partner name	Data Integration	Data governance and ethics	Compliance and regulations	Risk Management
Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency		LUT Business School	LUT Regulatory Frameworks	Consent and confidentiality; Bias representation; data quality and management

As it can be observed in Table 12, partners emphasize compliance with regulations, risk management, and the ethical use of data.

6.3.1 Data Integration and Governance

As far as Battery Ageing Lab Data and Fleet Data from Vehicles are concerned, partners such as AVL and OEMs, Tier-1 suppliers, and fleet operators focus on integrating data through engineering teams using FMU/FMI standards and APIs via standardized protocols like HTTP. This integration ensures seamless data flow and consistency across different platforms and systems, facilitating collaborative efforts and comprehensive data analysis.

For general standards, communication, and dissemination materials, all partners follow guidelines set by the European Federation for Welding, Joining and Cutting (EWF). This uniform approach ensures that data governance is consistent, promoting clarity and transparency in all communications and dissemination efforts.

Regarding research data, partners like FLE-FI ensure that data related to gear tester data, journal bearing tester data, flange connection tester data, FEA simulations, time-domain simulations, and test pictures are governed by their research engineers and project groups. This centralized governance guarantees that the data remains secured, up-to-date, and valid throughout the project lifecycle.

6.3.2 Compliance and Regulations

Regarding Battery Ageing Lab Data and Fleet Data from Vehicles, compliance with automotive standards and horizontal regulations is paramount for AVL and fleet operators. They adhere to the EU battery regulation, Data Act, Data Governance Act, and AI Act, ensuring that all activities align with the stringent requirements set by these frameworks.

For research-related data, FLE-FI follows internal company policies and software license agreements, ensuring that all data handling practices comply with relevant regulations and policies. Moreover, Ikerlan and LUT Business School adhere to their respective regulatory frameworks, ensuring that data handling practices meet organizational standards. For LUT, this includes managing consent and confidentiality,

addressing bias representation, and maintaining high data quality and management standards.

6.3.3 Risk Management

The primary risk identified by AVL and fleet operators is the potential for revealing intellectual property (IP). In the B2B fleet context, privacy is less of a concern compared to the risk of exposing proprietary information. These partners implement stringent measures to mitigate this risk and protect their valuable IP.

For FLE-FI, the focus is on ensuring data security, validity, and currency. Their risk management strategies revolve around keeping data secure and ensuring it remains relevant and accurate for research purposes.

Furthermore, LUT Business School and external research agencies focus on managing risks related to consent and confidentiality, bias representation, and data quality. Their comprehensive risk management framework addresses these potential issues, ensuring ethical and effective data use.

The R3-Mydas project demonstrates a robust commitment to data governance and ethics. Through careful integration, strict compliance with regulations, and comprehensive risk management strategies, partners like AVL, OEMs, FLE-FI, Ikerlan, and LUT Business School ensure that data is handled responsibly and ethically. This approach not only safeguards the integrity and security of the data but also aligns with the broader objectives of transparency, compliance, and innovation within the project.

6.4 Monitoring and Evaluation plan

This section describes the plan for monitoring and evaluating the success of the data management activities. In the R3-Mydas project, the Monitoring and Evaluation Plan is essential to ensure the effectiveness and quality of various datasets and activities undertaken by the partners. This plan focuses on the accuracy and reproducibility of models and stakeholder engagement, ensuring that the project's outputs are valuable and relevant to all stakeholders involved.

Table 13: Monitoring and Evaluation Plan.

Sequential number	Name of the dataset	Partner name	Monitoring and evaluation	Stakeholder engagement
1.	Battery ageing lab data	AVL	Reproducibility of models	internal: skill teams; customers: OEM-s, Tier-1 - suppliers
2.	Fleet data from vehicles	OEM-s; Tier-1-s; fleet operators	Accuracy and reproducibility of models	internal: skill teams; customers: OEM-s, Tier-1 - suppliers
3.	Tasks descriptions	DBL, TMCOMAS, AVL, FLENDER	NA	Demo cases partners

Sequential number	Name of the dataset	Partner name	Monitoring and evaluation	Stakeholder engagement
4.	Workstations analysis	DBL, TMCOMAS, AVL, FLENDER		
5.	Users profiles	DBL, TMCOMAS, AVL, FLENDER		
6.	Ergonomics data	DBL, TMCOMAS, AVL, FLENDER		
7.	Neurophysiological data (eye tracking)	DBL, TMCOMAS, AVL, FLENDER		Demo cases partners, users involved in experiments can request data to be deleted according to the informed consent
8.	Neurophysiological data (EEG)	DBL, TMCOMAS, AVL, FLENDER		
9.	Neurophysiological data (galvanic)	DBL, TMCOMAS, AVL, FLENDER		
10.	Microsoft Office	EITM Internal access	through the guidelines on FAIR Data Management in Horizon Europe	All the project partners
11.	Microsoft Form	ALL		
12.	Standards	All	NA	Local/Regional/National authorities
13.	Communication and dissemination material	All	Through the KPI's previously developed: Project Website; Social Media accounts	EU manufacturing companies, SMEs and MID-CAPs
14.	Stakeholder Analysis and community building strategy	All	Through the KPI's previously developed: Interactive workshops; Open day events	EU manufacturing companies, SMEs and MID-CAPs
15.	Gear tester data	FLE FI	Review of final documentation	FLE FI, company policy
16.	Journal bearing tester data	FLE FI		
17.	Flange connection tester data	FLE FI		
18.	FEA simulations	FLE FI		
19.	Time-domain simulations	FLE FI		
20.	Test pictures	FLE FI		
21.	Responses to user acceptance questionnaires	HUA	T5.3 leadership and WP5 leadership	Industrial community
22.	Residual stresses for induction hardened gear	Ikerlan	N/A	Flender, no concerns identified
23.	Lithium-ion battery aging dataset based on electric vehicle real-driving profiles	ITML	A quality metric to quantify and monitor certain categories of errors (missing,	the EV batteries manufacturers and the relevant automotive industries
24.	EVBattery: Large-scale EV dataset for	ITML		

Sequential number	Name of the dataset	Partner name	Monitoring and evaluation	Stakeholder engagement
	battery health and capacity estimation.		duplicates, incomplete, etc)	
25.	SICWELL Dataset	ITML		
26.	Minutes of the meetings	All	issues with data versioning or concurrent access.	All partners
27.	Gearbox simulation models	LUT, Flender	Accessibility and confirmation of data removal.	LUT, Flender
28.	Gearbox models	LUT, Flender		
29.	Measurement data	LUT, Flender	Accessibility and confirmation of data removal	LUT, Flender
30.	Experiment protocols	LUT, Flender		
31.	Scientific papers in production	LUT, Flender		
32.	Impact of remanufacturing process on industrial logistics	LUT Business School & external research agency	Regular review and audit of data; seek, document and act appropriately upon feedback; Evaluate the appropriateness of data management. Metrics are to be determined.	Project team members.
33.	Real crankshaft dimensions (geometric data)	TMCOMAS	N/A	TMCOMAS, ZIKNES and AIMEN
34.	Robot program for scanning the part	AIMEN/TMCOMAS		
35.	cloud point data (scanning the crankshaft)	AIMEN/TMCOMAS		
36.	laser cladding parameters	AIMEN/TMCOMAS		
37.	Robot program to weld the part	AIMEN/ZIKNES		
38.	welding quality monitoring data	AIMEN/TMCOMAS		
39.	Robot program for scanning the part after the welding	AIMEN/TMCOMAS/ZIKNES		
40.	final validation of the MYDAS solution	TMCOMAS		

6.4.1 Monitoring and Evaluation

Regarding Battery Ageing Lab Data and Fleet Data from Vehicles, AVL and OEMs, Tier-1 suppliers, and fleet operators emphasize the reproducibility and accuracy of their

models. AVL aims to reproduce the models developed from battery ageing lab data, while fleet operators focus on ensuring the models' accuracy and reproducibility using fleet data. These evaluations are crucial for maintaining high standards in data integrity and model performance.

As far as Task Descriptions, Workstations Analysis, Users Profiles, and Ergonomics Data are concerned, partners DBL, TMCOMAS, AVL, and FLENDER are responsible for monitoring task descriptions, workstation analyses, user profiles, and ergonomics data. Although specific monitoring methods are not detailed, these datasets are critical for understanding and improving the workflows and ergonomics within the project.

For datasets involving neurophysiological data such as eye tracking, EEG, and galvanic responses, DBL, TMCOMAS, AVL, and FLENDER ensure that the data is monitored through demo cases and partners. Users involved in experiments can request data deletion according to informed consent, highlighting the project's commitment to ethical considerations and user privacy.

Moreover, EITM and all partners utilize Microsoft Office and Forms for data management, following the FAIR Data Management guidelines in Horizon Europe. This ensures that data is findable, accessible, interoperable, and reusable, facilitating efficient project management and collaboration.

In addition to this, all partners monitor communication and dissemination materials using predefined KPIs, such as the project website and social media accounts. The standards are governed by local, regional, and national authorities, ensuring compliance and widespread engagement.

FLE-FI monitors gear tester and journal bearing tester data by reviewing final documentation, ensuring data integrity and relevance. This approach is consistent with company policies and aims to maintain high-quality standards. ITML monitors the lithium-ion battery aging dataset by applying quality metrics to quantify and monitor errors such as missing data, duplicates, and incomplete data. This meticulous approach ensures the reliability and accuracy of the dataset, which is vital for the automotive industries and EV battery manufacturers.

Regarding, Impact of Remanufacturing Process on Industrial Logistics is concerned, LUT Business School and an external research agency regularly review, and audit data related to the impact of remanufacturing processes on industrial logistics. They seek and document feedback, evaluate data management appropriateness, and determine metrics to ensure continuous improvement.

Finally, it is worth mentioning that all partners monitor meeting minutes to address issues with data versioning or concurrent access. Similarly, LUT and Flender focus on the accessibility and confirmation of data removal for gearbox simulation models, ensuring data integrity and proper management.

6.4.2 Stakeholder Engagement

The internal skill teams and external customers, including OEMs and Tier-1 suppliers, are the primary stakeholders engaged in evaluating battery ageing lab data and fleet data. This collaboration ensures that the data and models developed are aligned with industry needs and expectations.

For neurophysiological data, stakeholders include demo case partners and users involved in experiments. This engagement ensures that the data collected is relevant and ethically managed, considering user consent and privacy.

EU manufacturing companies, SMEs, and MID-CAPs are the primary stakeholders for communication and dissemination materials. These stakeholders are engaged through interactive workshops, open day events, and digital platforms, ensuring widespread dissemination and impact.

Furthermore, project team members are the key stakeholders for data related to the impact of remanufacturing processes. Their engagement ensures that the data management practices are appropriate and continuously improved based on feedback and audits.

Finally, stakeholders for Residual Stresses for Induction Hardened Gear dataset include Flender, with no specific concerns identified, ensuring straightforward data management and evaluation.

Overall, the Monitoring and Evaluation Plan in the R3-Mydas project emphasizes accuracy, reproducibility, and stakeholder engagement. By involving relevant stakeholders and adhering to high standards of data management, the project ensures that its outputs are valuable, reliable, and ethically managed.

6.4.3 Data Management Plan in R3-Mydas Work Packages

Data management within the R3-Mydas project is overseen by Task 1.3 "Data Management" [M1-M36], led by NCI. This task establishes the guidelines for technically sound, legally compliant, and ethically appropriate data collection, usage, and sharing among consortium partners, covering all data sets created and utilized to achieve the project's objectives throughout the data lifecycle as implemented in the R3-Mydas use cases. The Data Management Plan (DMP) will detail the relevant technical and organizational policies, procedures, safeguards, and controls to ensure legal compliance and ethical alignment within the context of the use cases. Additionally, the plan aims to facilitate legitimate downstream uses of personal data in accordance with the GDPR and the Data Governance Act, as well as non-personal data under Regulation 2018/1807 on the free flow of non-personal data and the Data Governance Act.

To ensure adherence to the outlined data management decisions in relation to the DMP, the following measures will be implemented in R3-Mydas:

- Work Package (WP) leaders are responsible for complying with the DMP specifications within their respective WPs.
- Each organization's project manager will oversee the DMP actions and will be available to the partner team for any DMP-related issues.
- Each organization's project manager must ensure that all personnel working on the project have read and will adhere to the R3-Mydas DMP principles.
- Data owners hold the ultimate responsibility for complying with the specifics of the R3-Mydas DMP and the related GDPR policies.
- For the overall project management activities, NCI, as the Project Coordinator, will ensure compliance with the DMP.

7 Conclusion

This Data Management Plan (DMP) underpins the management procedures and policies of the R3-Mydas project, aligning with the objectives of Work Package 1 (WP1: Project Management & Coordination). At the outset, the document provides a detailed description of the various data types to be utilized across all project activities, offering a comprehensive view of the R3-Mydas data lifecycle and categorizing different data assets.

The DMP outlines the project's policies and adherence to the General Data Protection Regulation (GDPR), ensuring that data privacy and security are prioritized. It specifies the potential data that may be generated or collected throughout the project's duration and details the strategies for making this data FAIR (Findable, Accessible, Interoperable, and Reusable). The plan also emphasizes maintaining the integrity and confidentiality of personal and sensitive data.

As a dynamic and evolving document, the DMP will be regularly updated to reflect new findings and refinements of data generated by the project. These updates will include the latest information on data types, actual data shared, metadata provided, and details on public sharing and relevant platforms. By continuously refining this document, R3-Mydas ensures robust and adaptable data management practices that support the project's goals of fostering sustainable circular value chains and advancing remanufacturing processes.

Appendix A The Data Management Survey

To ensure comprehensive data management throughout the R3-Mydas project, a detailed data survey was conducted with all project partners using an Excel file. This survey aimed to gather essential information on the various datasets to be used, generated, and managed during the project. The data survey encompassed multiple aspects of data management, which are crucial for maintaining data integrity, security, and compliance with relevant regulations. Specifically, the Excel file contained the following columns, each designed to capture specific aspects of data management:

- ✓ **#**: The sequential number for tracking purposes.
- ✓ **Name of the dataset**: The name of the dataset to be used in the project.
- ✓ **Dataset description**: Describe the utility of the dataset in the context of the project.
- ✓ **Partner name**: The name of the partner contributing to the data.
- ✓ **Responsibilities and roles**: Who is responsible for managing the data within each partner organization, and what are their specific roles and responsibilities?
- ✓ **Task no.**: Identification number or code for the specific task related to the data.
- ✓ **Work package no.**: Identification number or code for the specific work package related to the data.
- ✓ **Data source**: Where will the data come from? Will it be collected through surveys, interviews, experiments, or other methods (instruments, software, etc.)?
- ✓ **Data type**: What kind of data will be collected? Will it be quantitative or qualitative? Will it be structured or unstructured?
- ✓ **Data format**: In which format will the data be stored (e.g., CSV, Excel, database, etc.)?
- ✓ **Data volume**: The expected size or volume of the data generated or managed by each partner (e.g., in megabytes, gigabytes, terabytes) based on similar past experiences.
- ✓ **Data velocity**: The expected data velocity (e.g., kilobytes/megabytes generated per week/month, etc.) based on similar past experiences.
- ✓ **Data availability**: When will the first set of data be available for the project?
- ✓ **Data collection frequency**: How often will the data be collected? Will it be collected at specific intervals or on an ongoing basis?
- ✓ **Data storage location**: Where will the data be stored? Will it be stored locally or remotely?
- ✓ **Data backup**: Outline the backup procedures to ensure data integrity and prevent loss.
- ✓ **Data access and sharing**: Who will have access to the data, and under what conditions? How will the data be shared with other partners in the project?
- ✓ **Data sharing agreement**: Is there a need for documentation of any formal agreements or contracts related to data sharing among project partners?
- ✓ **Data security and privacy**: What measures will be taken to ensure the security and privacy of the data? Outline measures to anonymize or pseudonymize personal data

if applicable. Will any sensitive information be collected or stored? Ensure compliance with GDPR and other relevant data protection regulations.

- ✓ **Data quality control**: How will the quality of the data be ensured? Will there be any data cleaning or validation processes in place?
- ✓ **Data analysis and reporting**: How will the data be analyzed, and what reports will be generated? Who will be responsible for this work?
- ✓ **Data retention and destruction**: How long will the data be retained, and under what conditions will it be destroyed?
- ✓ **Data documentation and metadata**: What procedures will be followed in the project for documenting data, including file naming conventions, folder structures, and accompanying documentation? How will metadata be managed and made accessible to facilitate data sharing and reuse?
- ✓ **Data versioning**: How will you manage multiple data versions?
- ✓ **Blockchain technology**: If the data are to be stored in a blockchain, please define the subset of data needed to be considered. [IF APPLICABLE]
- ✓ **Digital product passport**: Are the data related to a digital product passport? Which subset of data is to be used? Who will have access to this information, and under what conditions? [IF APPLICABLE]
- ✓ **International data spaces (IDS)**: Are the data related to IDS? How will these data be used? Will any sensitive information be shared through these spaces? [IF APPLICABLE]
- ✓ **Machine learning model name**: Name or identifier of the machine learning model being used for specific tasks. [IF APPLICABLE]
- ✓ **Machine learning description**: What type of machine learning model will be developed, and how will it be trained and validated? Who will be responsible for this work? [IF APPLICABLE]
- ✓ **Model development and training**: What type of version control mechanisms will be implemented to track changes in datasets and ensure reproducibility? Which protocols will be established for organizing and cataloging data to facilitate efficient retrieval and reuse? Is there a need for labeling and annotating data processes? [IF APPLICABLE]
- ✓ **Model deployment and inference**: How will trained ML models be deployed in "production" environments considering scalability, latency, and resource constraints? What type of monitoring and feedback loops will be implemented to assess model performance and enable model iteration? [IF APPLICABLE]
- ✓ **Model maintenance**: What procedures are needed for maintaining and updating the machine learning model over time? [IF APPLICABLE]
- ✓ **Data integration**: How will data from different sources (e.g., blockchain, digital product passport, IDS) be integrated into the overall data management plan? What standards or protocols will be used to ensure interoperability? [IF APPLICABLE]
- ✓ **Data governance and ethics**: Who will be responsible for ensuring that the data is managed in an ethical and transparent manner? Will there be any specific guidelines or principles governing the use of these data? How will potential biases in the data be assessed to mitigate their impact on ML models?
- ✓ **Compliance and regulations**: What regulatory frameworks will apply to these data, and how will they be addressed? Will any specific licenses or permits be required?
- ✓ **Risk management**: What are the potential risks associated with using these data, and how will they be mitigated?

- ✓ **Monitoring and evaluation**: How will the success of the data management plan be monitored and evaluated? What metrics will be used to assess its effectiveness?
- ✓ **Stakeholder engagement**: Who are the key stakeholders for these data, and how will their needs and concerns be addressed throughout the data management process?

Firstly, partners were asked to provide a sequential number for tracking purposes, along with the name and a detailed description of each dataset to be used within the project. This initial information helps in understanding the utility and context of the dataset within the project's scope.

Next, partners were required to identify their organization and specify the roles and responsibilities of individuals managing the data. This includes detailing the responsibilities for data handling and the specific tasks and work packages related to the data.

The survey collected information on the data sources, types, and formats. Partners were asked to describe where the data would originate, whether it would be collected through surveys, experiments, or other methods, and to specify the nature of the data, such as whether it is quantitative or qualitative, structured or unstructured. They were also asked to indicate the format in which the data would be stored, such as CSV, Excel, or database formats.

To plan for storage needs, partners provided estimates of the data volume and velocity, describing the expected size of the data and the frequency at which it would be collected. This information helps in anticipating storage requirements and ensuring adequate infrastructure is in place.

Data availability and collection frequency were also covered, detailing when the first set of data would be available and how often data would be collected. Partners specified the locations where data would be stored, whether locally or remotely, and described their data backup procedures to ensure data integrity and prevent loss.

Regarding data access and sharing, partners indicated who would have access to the data and under what conditions, as well as any formal agreements or contracts related to data sharing. They also outlined the measures to be taken for data security and privacy, ensuring compliance with GDPR and other relevant data protection regulations.

Moreover, the survey addressed data quality control measures, asking partners to describe any data cleaning or validation processes in place. It also inquired about how data would be analysed and reported, including the roles responsible for these tasks.

Data retention and destruction policies were detailed, specifying how long the data would be kept and under what conditions it would be destroyed. Partners provided information on data documentation and metadata, including procedures for file naming, folder structures, and metadata management to facilitate data sharing and reuse.

For data versioning, partners explained how they would manage multiple versions of data. Additionally, the survey included questions on the use of blockchain technology, digital product passports, and International Data Spaces (IDS), if applicable, to understand how these technologies would be integrated and managed.

Machine learning aspects were covered, asking partners to provide the names and descriptions of any machine learning models used, and to describe the development, training, deployment, and maintenance processes for these models. They also explained how data from different sources would be integrated and the standards or protocols to be used to ensure interoperability.

The survey also gathered information on data governance and ethics, including the principles and guidelines governing data use, and how potential biases in data would be assessed and mitigated. Compliance with regulatory frameworks was addressed, with partners identifying any specific licenses or permits required.

Risk management strategies were also included, asking partners to identify potential risks associated with data use and how these risks would be mitigated. Finally, the survey covered monitoring and evaluation plans, detailing how the success of the data management plan would be assessed, and stakeholder engagement, explaining how the needs and concerns of key stakeholders would be addressed throughout the data management process.

This thorough data survey ensures that all aspects of data management are well-planned and executed, supporting the R3-Mydas project's objectives and compliance with ethical and legal standards.