This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101135988

AI-Enabled Data Lifecycles Optimization and Data Spaces Integration for Increased Efficiency and Interoperability

Project acronym:	PLIADES
Project name:	AI-Enabled Data Lifecycles Optimization and Data Spaces Integration for Increased Efficiency and Interoperability
Grant Agreement No:	101135988
Call:	HORIZON-CL4-2023-DATA-01
Topic	HORIZON-CL4-2023-DATA-01-02
Type of action:	HORIZON Research and Innovation Actions
Start of project:	01/01/2024

Deliverable 2.2

Users Requirements and Use Cases

Work Package:	**WP2 – User Requirements and Specifications**
Task:	**T2.2**
Lead Beneficiary:	**INNOVALIA**
Due Date:	**31/12/2024**
Submission Date:	**31/01/2025**
Deliverable Status:	**Draft**
Deliverable Type:	**Report**
Dissemination Level:	**Public**

Authors

Surname	First Name	Beneficiary
Rodriguez	Silvia	INNOVALIA
Wurzinger	Lisa	AVL
Maria	Papaspyropoulou	ATLA
Tsiakas	Kosmas	CERTH
Mariolis	Ioannis	CERTH
Nasoulis	Christos	CERTH
Tsamis	George	CERTH
Giakoumis	Dimitrios	CERTH
Alexiou	Dimitrios	CERTH
Kontodina	Theodora	CERTH
Kirstein	Franziska	BOR
-	-	VICOM, Taltech, DENN, CVUT, PATRIC, ZERO, MU-EPS, CEIT, IDSA, TNO, CIC BioGUNE, BCCAM, WR, Libattion, SIPBB, I4BD, EURECAT, THI, UC3M [1]

Reviewers

Surname	First Name	Beneficiary
Insabato	Marco Antonio	E@W
Jirovský	Václav	PATRIC

Version History

Version	Date	Modifications made by
0.1	10/11/2024	INNOVALIA
0.2	25/11/2024	INNOVALIA
0.3	13/12/2024	INNOVALIA
0.4	18/12/2024	INNOVALIA
0.5	08/01/2025	INNOVALIA (after reviewers' feedback)
1.0	31/01/2025	INNOVALIA

[1] The listed partners provided their input indirectly through circulated templates processed by INNOVALIA that authored the corresponding sections of the deliverable based on the consolidated feedback.

Disclaimer

Executive Summary

The goal of D2.2 “User Requirements and Use Cases” is to provide a thorough overview of the derived user requirements and define the specifications of the PLIADES Use Cases. The present deliverable reports on the activities developed throughout T2.2 “Consolidation of user and system requirements”, as well as T2.4 “Design of the Use Cases for the deployment of the multiple project data spaces”. Through these tasks, the user and system requirements of the PLIADES Use Cases have been defined, resulting in clear requirements that will lead to the demonstration of the PLIADES tools in real-world scenarios.

In order to achieve this, a methodology for requirements engineering was adopted, based on standards which include discrete and specific steps. First of all, the relevant stakeholders were identified, as the users of the PLIADES platform for the different domains. Business requirements were also created, in order to collect specific use case information and requirements from each partner. These surveys were filled in by the use case-involved partners, leading to a detailed analysis of the use case information, their needs and requirements. Finally, all requirements were further analysed and processed, in order to create the final, consolidated set of requirements that represent all PLIADES use cases without any overlaps or conflicts. Additionally, all requirements have been sorted according to their prioritization, as has been reported by the users.

The PLIADES project will be demonstrated in six use cases focusing on direct advancements in key AI technologies, oriented around various data spaces, such as Mobility, Healthcare, Industrial, Energy and Green Deal. This document provides a clear description of each use case, starting with the description of their concept, the main objectives and the involved participants. For each use case, specific scenarios were identified, describing both their current and their future processes. For each scenario, requirements have been collected, allowing the mapping of use case needs with the PLIADES specifications. Starting from high-level use case applications, each use case was divided into “sub-use cases”, which were then analyzed in detail to define the corresponding functional and non-functional requirements, along with their priority level.

The document concludes with the future steps for the project, which are directly related to the requirements that have been extracted.

Through the above, D2.2 is a first and important step in system functionality preparation. It serves as the background for the establishment of the User-Centred-Design (UCD) oriented research and development efforts that are foreseen in the further steps of the project. More specifically, UCD was used to identify the PLIADES system from a technological point of view in order to fully address task priority, driving the system design from a user-centric perspective facilitating the design and development of the system architecture in task T2.3 “Design and architecture of the platform and data integration strategies”.

Table of Contents

Executive Summary ... 3
Table of Contents ... 4
List of Figures ... 7
List of Tables ... 8
List of terms and definitions ... 9
1 Introduction ... 11
1.1 Scope of the deliverable ... 11
1.2 Relation to other activities and deliverables ... 11
1.3 Structure of the deliverable ... 11
2 Requirements investigation approach ... 12
2.1 Methodology ... 12
2.2 Role of users in the PLIADES framework ... 15
3 Use Cases ... 17
3.1 Common barriers across use cases ... 17
3.1.1 Legal Frameworks ... 17
3.1.2 Social Frameworks ... 18
3.2 Use Case 1: Integrating Data Lifecycles of sustainability, operations and process industry manufacturing operations ... 19
3.2.1 Use case overview ... 19
3.2.2 Background ... 19
3.2.3 Future scenario ... 23
3.2.4 Objectives ... 25
3.2.5 Actors involved ... 26
3.2.6 User & system requirements ... 27
3.2.7 Expected benefits and KPIs ... 29
3.3 Use Case 2: Integrating data life cycles of service robot to improve HRI with end users ... 30
3.3.1 Use case overview ... 30
3.3.2 Background ... 31
3.3.3 Future scenario ... 36
3.3.4 Objectives ... 38
3.3.5 Actors involved ... 39
3.3.6 User & system requirements ... 40
3.3.7 Expected benefits and KPIs ... 41

3.4 Use Case 3: Integrating data lifecycles of personalized medicine services to improve diagnostic and prognostic clinical prediction models 42
3.4.1 Use case overview 42
3.4.2 Background 42
3.4.3 Future scenario 46
3.4.4 Objectives 50
3.4.5 Actors involved 51
3.4.6 User & system requirements 52
3.4.7 Expected benefits and KPIs 52
3.5 Use Case 4: Integrating data life cycles of smart vehicles for CCAM operations and ADAS/AD functions 54
3.5.1 Use case overview 54
3.5.2 Background 55
3.5.3 Future Scenario 61
3.5.4 Objectives 65
3.5.5 Actors involved 66
3.5.6 User & system requirements 67
3.5.7 Expected benefits and KPIs 69
3.6 Use Case 5: Integrating data life cycles of WEEE/batteries management and car parts manufacturing operations 70
3.6.1 Use case overview 70
3.6.2 Background 71
3.6.3 Future scenario 73
3.6.4 Objectives 75
3.6.5 Actors involved 76
3.6.6 User & system requirements 76
3.6.7 Expected benefits and KPIs 77
3.7 Use Case 6: Integrating professional service robot data life cycles to improve HRI with robot operators 77
3.7.1 Use case overview 77
3.7.2 Background 78
3.7.3 Future scenario 84
3.7.4 Objectives 87
3.7.5 Actors involved 87
3.7.6 User & system requirements 89

3.7.7 Expected benefits and KPIs....89
3.8 Definition of PLIADES use cases....91
4 Requirements and analysis....92
5 Conclusions and next steps....110
6 References....112
7 ANNEX I: Trial Handbook Chapter 2 / Business Requirements Questionnaire....113
Preface....113
1 Use Case Framework and Barriers....114
1.1 Legal Framework....114
1.2 Social Framework....117
1.3 Economic Framework....117
2 Business Scenario....118
2.1 Current Business Scenario....118
2.2 Future Business Scenario....119
2.3 Business Process....119
2.3.1 Business Process In Detail....119
2.3.2 Business Objectives....119
3 BUSINESS INDICATORS....122
4 ACTORS INVOLVED....123
5 EXPECTED BENEFITS....124
6 BUSINESS REQUIREMENTS....126

List of Figures

Figure 1 ToC for the Trial Handbook Chapter 2 / Business Requirements Questionnaire 14
Figure 2 DENNDATA – Product digital platform's scheme 21
Figure 3 Current manual process gait training with PTR robot 32
Figure 4 Current robots from CERTH performing human action recognition & receptionist tasks 33
Figure 5 Current platform architecture of CIC BioGUNE 44
Figure 6 Data lifecycle 46
Figure 7 Data space-based architecture 47
Figure 8 Sharing data and models scenario 48
Figure 9 Collaborative data/service exchange and third-party data acquisition scenario 48
Figure 10 Future Business scenario for data related to Battery Operation 73
Figure 11 Future scenario for data related to Battery Digital Twin 73
Figure 12 GoBe Robot in a production setting today 79
Figure 13 Operator driving robot via joystick to destination 80
Figure 14 Robots from CERTH performing human action recognition & inspection tasks 81
Figure 15 Use Case 1 functional requirements categorized by type 97
Figure 16 Use case 2 and 6 functional requirements categorized by type 98
Figure 17 Use case 3 functional requirements categorized by type 98
Figure 18 Use Case 4 Functional Requirements categorized by type 99
Figure 19 Use Case 5 Functional requirements categorized by type 99
Figure 20 Bar plot for the complete set of functional requirements categorized by type 100
Figure 21 Use Case 1 non-functional requirements categorized by type 101
Figure 22 Use Case 2 and 6 non-functional requirements categorized by type 101
Figure 23 Use Case 3 non-functional requirements categorized by type 101
Figure 24 Use case 4 non-functional requirements categorized by type 102
Figure 25 Bar plot for the global Non-functional requirements categorized per type 102
Figure 26 Use Case 1 Functional Requirements categorized per area 103
Figure 27 Use case 2 and 6 Functional Requirements categorized per area 103
Figure 28 Use Case 3 Functional Requirements categorized per area 104
Figure 29 Use Case 4 Functional Requirements categorized per area 104
Figure 30 Use Case 5 Functional Requirements categorized by area 105
Figure 31 Use Case 1 non-functional requirements categorized per area 106
Figure 32 Use Case 2 and 6 non-functional requirements categorized per area 106
Figure 33 Use Case 3 non-functional requirements categorized per area 106
Figure 34 Use case 4 non-functional requirements categorized per area 107
Figure 35 Functional Requirements categorized per area. 107
Figure 36 Non-Functional Requirements categorized per area 108

List of Tables

Table 1 Definitions 9
Table 2 UC 1 business objectives 25
Table 3 UC1 actors involved 26
Table 4 UC1 requirements 27
Table 5 UC1 Benefits and KPIs 30
Table 6 UC 2 business objectives 38
Table 7 UC2 actors involved 39
Table 8 UC2 requirements 40
Table 9 UC2 benefits and KPIs 41
Table 10 UC 3 business objectives 50
Table 11 UC3 actors involved 51
Table 12 UC3 requirements 52
Table 13 UC3 benefits and KPIs 53
Table 14 UC 4 business objectives 65
Table 15 UC4 actors involved 66
Table 16 UC4 requirements 67
Table 17 UC4 benefits and KPIs 69
Table 18 UC 5 business objectives 75
Table 19 UC5 actors involved 76
Table 20 UC5 requirements 76
Table 21 UC5 benefits and KPIs 77
Table 22 UC 6 business objectives 87
Table 23 UC6 actors involved 87
Table 24 UC6 requirements 89
Table 25 UC6 Benefits and KPIs 90
Table 26 Summary of high-level use cases and their sub-use cases 91
Table 27 Functional user requirements of the PLIADES platform 92
Table 28 Non-functional user requirements of the PLIADES platform 95
Table 29- Sector/Industry Framework & Impact summary 115
Table 30- Technologies Framework & Impact summary 115
Table 31- Products/Services Framework & Impact summary 116
Table 32- Business Scenario 118
Table 33- Business Process: Business Objective Summary 120
Table 34- Business Process: Business Indicators Summary 122
Table 35- Business Process: Actors involved summary 123
Table 36- Business Process: Expected benefits summary 125
Table 37- Business Process: Business requirements summary 127

List of terms and definitions

Table 1 Definitions

Abbreviation	Definition
2D	Two-Dimensional
3D	Three-Dimensional
AD	Autonomous Driving
ADAS	Advanced Driver Assistance Systems
AI	Artificial Intelligence
API	Application Programming Interface
BESS	Battery Energy Storage Systems
BMS	Battery Management System
CAD	Computer-Aided Design
CCAM	Connected, Cooperative and Automated Mobility (CCAM)
CDR	Community Design Regulation
CE	Conformité Européenne, or European Conformity
COCO	Common Objects in Context
CoPa	Federal Act on Copyright and Related Rights
DGA	Data Governance Act
DL	Deep Learning
DS	Data Space
DSBA	Data Spaces Business Alliance
DSSC	Data Spaces Support Centre
ERS	European Research Survey
EU	European Union
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HRI	Human-Robot Interaction
ICT	Information Communications and Technology
IDSA	International Data Spaces Association
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
IP	Intellectual Property
ISMS	Information Security Management Systems

ISO	International Organization for Standardization
IT	Information Technology
IVDr	In Vitro Diagnostic Regulation
KPI	Key Performance Indicator
LiDAR	Light Detection and Ranging
MDR	Medical Device Regulation
ML	Machine Learning
NMR	Nuclear Magnetic Resonance
POC	Proof-of-Concept
QMS	Quality Management Systems
R&D	Research and Development
RGB	Red-Green-Blue
ROS	Robot Operating System
RSU	Road Side Unit
SME	Small Medium Enterprise
SUS	System Usability Scale
SoH	State of Health
SUS	System Usability Score
TDS	Test Data Space
TRIPS	Trade-Related Aspects of Intellectual Property Rights
UAV	Unmanned Aerial Vehicles
UCD	User-Centered Design
UR	User Requirement
V2X	Vehicle-to-Everything
WP	Work Package
YOLO	You Only Look Once

1 Introduction

1.1 Scope of the deliverable

The purpose of this deliverable (D2.2 User Requirements and Use Cases) is to clearly define the six PLIADES use cases that span across the five data spaces of the project identifying the main user needs for the implementation of the PLIADES framework. For each use case, intensive and extensive discussions have been carried out to outline the user requirements all through the different scenarios defined in the use cases. The current deliverable presents the current situation for the data processes of the use cases, their main barriers and challenges as well as the objectives lying under the adoption of Data Spaces throughout the data lifecycle operations. Moreover, the user needs per use case identified in this deliverable, both functional and non-functional, are presented and then analyzed to deliver a complete design of the use cases in relation with the technical specifications that can be found and are further explained in D2.3 "System Technical Specifications and PLIADES Framework Architecture".

1.2 Relation to other activities and deliverables

The current report describes the results from task T2.2 "Consolidation of user and system requirements" and task T2.4 "Design of the Use Cases for the development of the multiple project data spaces" of the PLIADES project, building upon other WP2 tasks, particularly Task T2.3 "Design and architecture of the platform and data integration strategies" where specific system requirements are analyzed and also, task T2.6 "Human factors across the data life cycle in multiple data spaces" where an specific analysis of relevant actors in the six use cases is being performed.

1.3 Structure of the deliverable

The deliverable is structured as reported below:

Section 1 introduces this report, by outlining the scope of the document, its relation to other activities and deliverables, and an overview of the structure of the deliverable.

Section 2 explains the methodology used for defining the user needs, requirements and KPIs and specifies the approach for the design of the future scenario in the six use cases. Moreover, a brief presentation of the different users and their roles in the use cases is provided.

Section 3 describes all the PLIADES use cases separately and contains their general concept description, the involved partners, the current and the future scenario, after the integration of PLIADES.

Section 4 presents the main outputs of the user requirements analysis and contains a detailed list of all the identified user requirements.

Section 5 concludes the document, summarizing the contributions of the WP2 activities carried out under T2.2 and T2.4 and presenting the next steps for the implementation of the PLIADES Use Cases.

Section 6 contains the references that have been used in this document.

Section 7 contains information regarding the business questionnaires that were used for the identification of the user requirements.

2 Requirements investigation approach

2.1 Methodology

The PLIADES project adopts a User-Centered Design (UCD) approach to ensure its framework addresses the diverse needs of its stakeholders across six distinct use case scenarios. To achieve this, the current deliverable employs a combined methodology to define the use cases, their future scenarios, and the requirements necessary for their realization. This approach integrates UCD principles with structured processes to systematically collect, analyse, and refine information across Work Package 2 (WP2), ensuring comprehensive and stakeholder-aligned requirements elicitation.

A key component of this methodology is the application of ISO/IEC/IEEE 29148:2018 Systems and software engineering — Life cycle processes — Requirements engineering, a globally recognized standard for requirements engineering. This standard provides a comprehensive framework for managing requirements as an interdisciplinary function, bridging the acquirer and supplier domains to establish and maintain requirements for systems, software, or services.

The process requires the following steps:

1. Stakeholder identification

The first step involves identifying stakeholders, as defined in the ISO/IEC/IEEE 29148:2018 standard. In the acquirer-supplier scenario, stakeholders are identified as the users or actors of the system. Stakeholder identification is a critical process as it ensures the inclusion of diverse perspectives, enabling the development of a framework that meets technical, operational, and contextual needs.

In the PLIADES project, stakeholders encompass a wide range of roles and responsibilities. Industrial players provide operational data and validate use case-specific requirements to ensure the framework's practical applicability in industrial environments. Technical teams contribute expertise in ICT (Information Communications and Technology), data science, and operational technology, designing and implementing robust, scalable solutions. Healthcare professionals and patients play an essential role in defining user-centric requirements, prioritizing privacy, usability, and ethical considerations specific to healthcare scenarios. Additionally, researchers and domain experts contribute theoretical and technical insights, driving innovation and ensuring the development of scientifically grounded solutions.

Each of the six use case scenarios requires either its own dedicated data space or multiple data spaces, tailored to meet the specific needs of these stakeholders.

The stakeholder identification process is carried out through structured mapping sessions in collaboration with project partners, in order to capture the aforementioned roles effectively. These sessions analyzed the roles, responsibilities, and expectations of each stakeholder group to ensure comprehensive representation. For example, each use case in PLIADES is executed by multidisciplinary teams, combining expertise in ICT, operational technology, and data science. This collaborative structure ensures that all technical, functional, and sector-specific perspectives are integrated into the project. These efforts are elaborated further in the “Actors Involved” subsections of each use case.

2. Collection of needs

The second step in the methodology focuses on systematic gathering and validating the needs of stakeholders to ensure the requirements elicitation process is robust, comprehensive, and scientifically grounded. This process was conducted for each use case in close collaboration with involved project partners, employing the structured framework of **Trial Handbook Chapter 2 –**

Business Requirements. This approach builds on methodologies carried out in FITMAN[2] and successfully applied in BOOST4.0.[3], QU4LITY[4] and XMANAI[5], demonstrating its proven effectiveness in large-scale, multidisciplinary settings.

For each use case, a **Business Requirements Questionnaire** was developed and completed collaboratively by the responsible teams. The questionnaire provided a standardized structure to ensure consistency across all six use cases while accommodating the unique characteristics and needs of each scenario. The Table of Contents (TOC) of this document is depicted in Figure 1. This standardized framework ensured:

- **Uniform Data Collection:** Providing a consistent methodology across use cases enabled comparative analysis and alignment with the overarching project objectives.
- **Customization:** While maintaining a common format, the questionnaire enabled the inclusion of sector-specific and use case-specific questions, addressing the unique challenges of diverse domains such as healthcare, mobility, industrial applications, etc.

The development and completion of the questionnaire involved multiple iterative review cycles and stakeholder engagement sessions to ensure the accuracy and relevance of the collected data. Several coordination meetings among PLIADES partners were held to:

- Analyze and refine the questions to ensure clarity and precision.
- Validate the completeness and alignment of the collected data with project objectives.
- Address feedback and evolving requirements as the understanding of each use case deepened.

The collected information was consolidated into a master document, serving as a centralized repository of validated needs and requirements. This document provides the foundation for:

- **Cross-Case Comparisons:** Allowing the identification of commonalities and differences among use cases to enhance interoperability and scalability across the PLIADES framework.
- **Traceability:** Linking each identified need to its corresponding requirement ensures clear accountability and alignment with the development process. This traceability fosters transparency and enables systematic updates as the project evolves.

The Business Questionnaire and its content is described in detail in Section 7 – ANNEX I: Trial Handbook Chapter 2 / Business Requirements Questionnaire.

[2] https://cordis.europa.eu/project/id/604674/reporting

[3] https://boost40.eu/

[4] https://qu4lity-project.eu/

[5] https://ai4manufacturing.eu/

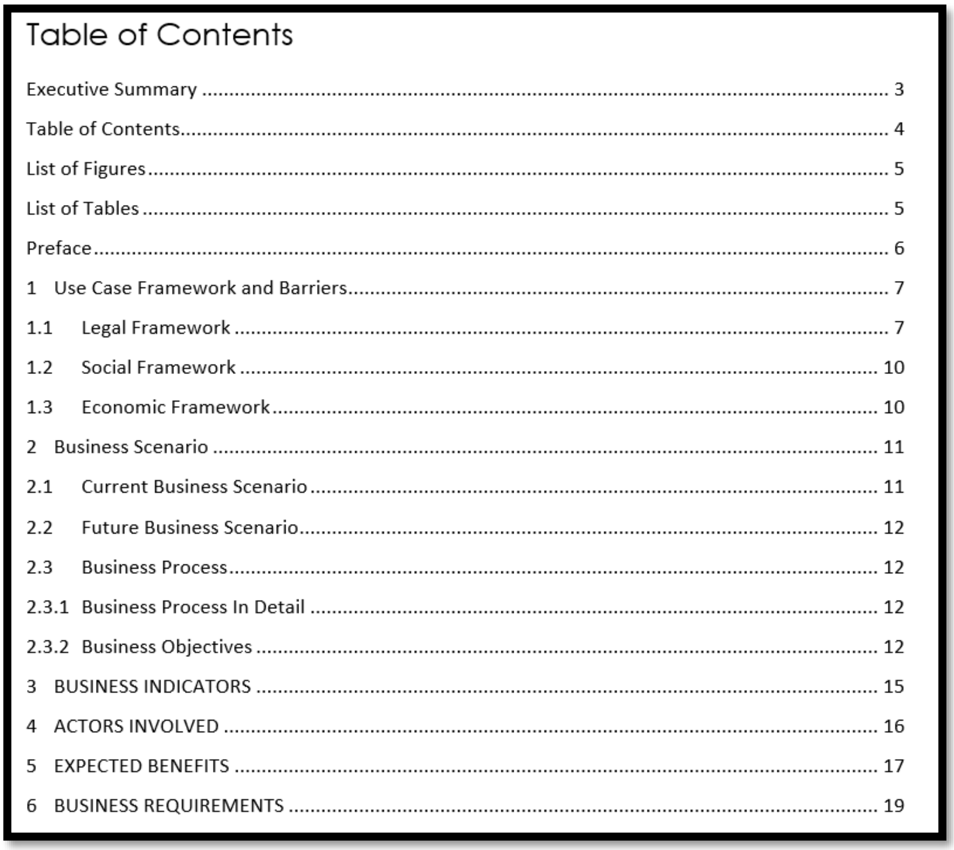
Table of Contents

Executive Summary 3
Table of Contents 4
List of Figures 5
List of Tables 5
Preface 6
1 Use Case Framework and Barriers 7
1.1 Legal Framework 7
1.2 Social Framework 10
1.3 Economic Framework 10
2 Business Scenario 11
2.1 Current Business Scenario 11
2.2 Future Business Scenario 12
2.3 Business Process 12
2.3.1 Business Process In Detail 12
2.3.2 Business Objectives 12
3 BUSINESS INDICATORS 15
4 ACTORS INVOLVED 16
5 EXPECTED BENEFITS 17
6 BUSINESS REQUIREMENTS 19

Figure 1 ToC for the Trial Handbook Chapter 2 / Business Requirements Questionnaire

3. Definition of requirements

The last step of the methodology emphasizes on the definition of specific requirements. The identification of specific needs through the questionnaire serves as the foundation for defining detailed requirements, which are formulated in accordance with the **ISO/IEC/IEEE 29148:2018** standard. This process ensures that the requirements are technically feasible, and aligned with stakeholder expectations.

In the final step, stakeholder needs are systematically translated into specific requirements, which are then subjected to a thorough analysis to ensure their quality and reliability. This analysis is conducted emphasizing in the following three key criteria:

- **Consistency:** the set of requirements contains individual requirements that are unique and free from conflicts or overlaps with other requirements in the set. Logical consistency is critical to avoid ambiguities that could complicate the implementation process. Consistency checks include cross-referencing related requirements to identify potential redundancies or contradictions.
- **Completeness:** the set of requirements stands alone, sufficiently describing the necessary capabilities, characteristics, constraints or quality factors required for the system. Completeness extends beyond core functionality to include critical aspects such as:
  - **Interoperability:** Ensuring seamless integration with other systems and platforms.
  - **Data security:** Addressing regulatory and operational requirements to protect sensitive information and privacy constraints.
  - **User experience:** Incorporating user-centric features that improve usability and accessibility.

  Any gaps in the requirements are addressed through iterative refinement processes, involving feedback loops with stakeholders, domain experts, and technical teams. These cycles ensure that no critical requirements are omitted and that the evolving needs of all stakeholders are incorporated.

- **Correctness:** Each requirement accurately reflects stakeholder needs and is technically feasible for implementation. Validation processes include:
  - **Empirical Validation:** Comparing the requirements against real-world scenarios and benchmarks to ensure relevance and applicability.
  - **Stakeholder Reviews:** Engaging stakeholders in regular review sessions to confirm that the requirements align with their expectations and objectives.

2.2 Role of users in the PLIADES framework

Users play a central role in the implementation of the PLIADES framework, as their needs and requirements drive the development and configuration of the platform. With 27 participants and 6 use cases, the PLIADES project necessitates careful consideration of the diverse needs of its different types of users, ranging from researchers and developers to operators and managers, each contributing unique perspectives and expertise to the framework's development and implementation. The role of users in the implementation of the framework is explained below:

- **Requirements Definition:** The users, through the different use cases, are the ones who define the functional and non-functional requirements of the framework. Their specific needs in each sector (industry, health, mobility, energy) guide the design of the platform, ensuring that it fits their requirements.
- **Participation in the Use Cases:** Some of the identified users will actively participate in the project execution, where the PLIADES framework is implemented and validated. Each use case represents a specific scenario with its own stakeholders and needs, allowing for a more precise adaptation of the framework.
- **Contribution of Knowledge and Experience**: Users contribute their knowledge and experience in their respective fields, which enriches the development of the framework. This knowledge is crucial to ensure that the implemented solutions are practical and effective. For example, annotators, who are part of the users, contribute their knowledge to refine the AI-generated annotations.
- **Role Specific:** Within the use cases, different user roles are identified, each with a specific type of impact:
  - **Data providers:** Responsible for data collection and data exchange.
  - **Data consumers:** Use data for AI model development, decision making and process improvement.
  - **ADAS (Advanced Driver Assistance Systems) system integrators:** Responsible for the integration of AI models into vehicular systems.
  - **Researchers:** Responsible for the integration, analysis and preparation of new data sources.
  - **IT (Information Technology) security specialists:** Implement and maintain security and privacy measures.
  - **Healthcare professionals:** Interact directly with robots and evaluate patient-related data.
  - **Financial administrators:** Manage the financial aspects of data integration and exchange processes.
  - **Project managers:** Oversee the integration of new data sources.
- **Platform Interaction:** Users interact with the PLIADES platform through user interfaces, data visualizations, and collaboration tools. The usability of these tools is crucial to ensure adoption and effectiveness of the framework.

- **Validation and Feedback:** Users validate the performance of the framework in real environments and provide feedback for continuous improvement. Their comments and suggestions are critical to identify areas of improvement and optimize platform performance. Users contribute to the definition of metrics to evaluate the quality of the annotations.
- **Platform Adoption:** User acceptance and adoption of the PLIADES framework is a critical factor for the success of the project. The platform must be easy to use, adaptable, and offer tangible benefits to users. Training and technical support are essential to facilitate user adoption of the platform.
- **Data governance:** Users are involved in establishing data governance policies to ensure quality, consistency, and regulatory compliance in the exchange and use of information.

In summary, users are the key stakeholders in the implementation of the PLIADES framework. Their active participation, knowledge and feedback are essential for the platform’s development, ensuring it remains useful, effective and adaptable to the diverse needs of the different sectors involved.

3 Use Cases

3.1 Common barriers across use cases

Each of the PLIADES use cases faces unique challenges, originating from the domain-specific requirements and needs. However, they also share some common aspects, such as the barriers to implement them, which are directly related to the legal frameworks and the social frameworks. The barriers which are common for all use cases are introduced and consolidated in this section. The barriers which are specific to each use case are introduced separately in the following sections.

3.1.1 Legal Frameworks

- General Data Protection Regulation (GDPR): This EU regulation ensures that personal data collected or processed through AI-driven systems and data integration technologies is handled securely and in compliance with privacy laws. Companies that fail to comply risk facing severe financial penalties and damage with their reputation. Compliance is further supported by the implementation of ISO/IEC 27001, which provides a secure framework for managing operational and personal data.
- EU Data Governance Act (DGA): Adopted in June 2022, it is designed to create a framework for the secure sharing and reuse of industrial and public sector data across the European Union. It aims to boost innovation by facilitating the exchange of data between businesses, individuals, and public institutions while maintaining strict safeguards for privacy and data security. The DGA introduces the concept of data intermediaries, trusted entities that manage data sharing without taking control of the data itself, promoting transparency and trust. Additionally, it supports data altruism, allowing individuals and companies to voluntarily share data for public good, such as in scientific research or public health. By enabling easier access to non-personal data, the DGA promotes collaboration and innovation across industries while ensuring compliance with EU privacy regulations.
- EU AI Act: It is poised to reshape the landscape of Industrial AI solutions, as it introduces a rigorous framework for the development, deployment, and maintenance of Artificial Intelligence systems in the sector. The regulation's emphasis on transparency, explainability, and safety will likely lead to the development of more robust and reliable AI-powered industrial solutions, such as predictive maintenance, quality control, and supply chain optimization. As a result, industries that adopt compliant AI solutions can expect improved efficiency, reduced downtime, and enhanced decision-making capabilities. Moreover, the AI Act's focus on human oversight and accountability will drive the development of more explainable and auditable AI models, ultimately building trust and confidence in the use of AI in industrial settings.
- Cybersecurity Act: This will have a profound impact on manufacturers of industrial machinery, as it introduces stricter requirements for the security of connected devices and the protection of sensitive data. With the increasing adoption of Industry 4.0 technologies, manufacturers are generating vast amounts of data from their machines, which must be stored, processed, and transmitted securely. The Cybersecurity Act's emphasis on secure-by-design principles and the use of secure protocols for data transmission will require manufacturers to rethink their data management strategies and invest in robust cybersecurity measures to protect their data spaces. This includes implementing secure data storage solutions, encrypting data in transit, and ensuring that data is properly segregated and access-controlled. By doing so, manufacturers can mitigate the risk of cyber-attacks, protect their intellectual property, and maintain the trust of their customers and stakeholders.
- Ethical Guidelines: Ethical concerns in AI and robotics include bias, transparency, and patient autonomy. Specifically, AI algorithms must be designed to avoid discrimination, and patients

must be fully informed about AI's role in their care. Human oversight is emphasized, ensuring AI assists rather than replaces medical judgement.

- Ethical Guidelines: Ethical concerns in AI and robotics include bias, transparency, and patient autonomy. Specifically, AI algorithms must be designed to avoid discrimination, and patients must be fully informed about AI's role in their care. Human oversight is emphasized, ensuring AI assists rather than replaces medical judgement.
- IP-rights: At the European level, IP rights for designs are regulated by the Community Design Regulation (CDR) and the Design Directive. These provide protection through both registered and unregistered designs across the EU, ensuring consistent rights within member states. In Switzerland, intellectual property (IP) rights for design data are governed by the Federal Act on Copyright and Related Rights (CopA) and the Federal Act on the Protection of Designs (Designs Act). These laws protect the creators' rights to their designs, including digital design data, against unauthorized use or reproduction. Internationally, design protection is covered by agreements such as the Berne Convention, the TRIPS Agreement (Trade-Related Aspects of Intellectual Property Rights), and the Hague Agreement on the International Registration of Industrial Designs. These frameworks establish minimum standards for design protection and allow for the registration of designs across multiple countries.
- ISO/IEC 27001, Information Security Management: ISO/IEC 27001 is a globally recognized standard for Information Security Management Systems (ISMS). It is essential for organizations dealing with large volumes of data, particularly sensitive and personal data, as it provides a framework to manage data security and ensure compliance with GDPR and other data protection regulations.
- ISO 9001: As the global standard for quality management systems (QMS). It applies to organizations involved in the design, development, and deployment of AI systems in mobility and industrial sectors, ensuring that products meet consistent quality standards.

3.1.2 Social Frameworks

- Privacy: Maintaining the privacy of subjects from whom data is obtained for the development of precision medicine algorithms is a top priority. Therefore, it is compulsory to comply with the GDPR and adhere to best practices to protect subjects' data privacy. Person anonymity is essential, which involves anonymizing records and implementing measures to prevent the identification of subjects from the collected data. Also, measures to avoid data manipulation and data hacking should be considered due to the sensible data used. This focus on privacy stems from a growing societal demand for transparency in how data is used and the ethical considerations of data ownership, especially as industries rely more heavily on AI for decision-making.
- Security: Cybersecurity is a significant social issue as more industries adopt AI and data-driven systems to optimize their operations. Within the context of the trial, securing the data lifecycle from potential breaches or cyberattacks is a top priority. The rise of cyber threats has made data security a central concern not only for businesses but for society at large, as the integrity of industrial data is crucial for ensuring trust in AI systems. The ISO/IEC 27001:2023 certification assists to maintain high security standards, a growing societal expectation in today's interconnected world. Cybersecurity measures, such as encryption, access control, and regular audits, must be embedded into all technology layers.
- Globalization: The trials also address challenges related to globalization and the cross-border sharing of data. The impact of this is seen in the complexity of complying with varying international regulations, such as the EU Data Governance Act, which seeks to ensure that data is shared securely and transparently across borders. Globalization drives the need for such frameworks, as industries increasingly collaborate internationally, but must manage the societal concerns around data sovereignty and regional privacy standards.

- Health and safety: AI-driven enhancements, such as improved traffic management and advanced driver-assistance systems, have the potential to reduce accidents, lower emissions, and enhance public health. Prioritizing health and safety metrics during development ensures that societal goals align with technical advancements.

3.2 Use Case 1: Integrating Data Lifecycles of sustainability, operations and process industry manufacturing operations

3.2.1 Use case overview

Use Case 1 will be implemented at DENN's facilities in Polinyà, Spain, leveraging DENNDATA, the company's digital platform, which enables real-time data collection, analysis, and optimization. The goal is to demonstrate the potential of data-driven decision-making in enhancing productivity, reducing environmental impact, and fostering interoperability within the industrial sector. Additionally, the use case will address critical aspects of data privacy and cybersecurity, ensuring secure data handling and compliance with industry standards. This integration aligns with the principles of Industry 4.0 and aims to validate the PLIADES platform through a robust industrial demonstrator.

3.2.2 Background

3.2.2.1 Participants

DENN, also known as Industrias Puigjaner S.A., is a Spanish company with extensive experience in the design and production of metal forming machinery. Its expertise encompasses processes such as spinning, flow forming, and rotary forging. DENN's machinery is primarily used in industries like automotive, aerospace, and ventilation, with a significant focus on precision engineering and robust operational performance. The company exports approximately 95% of its production, reflecting its capacity to meet diverse international manufacturing standards.

Within the PLIADES framework, DENN is focusing on adapting its DENNDATA platform to align with complex multi-data space requirements. This includes integrating data lifecycles related to sustainability and operational processes in the manufacturing sector. Initial phases have highlighted DENN's ability to address interoperability challenges and implement scalable solutions for data management in industrial environments.

TECNALIA is the largest centre of applied research and technological development in Spain, a benchmark in Europe and a member of the Basque Research and Technology Alliance. It collaborates with companies and institutions to improve their competitiveness, people's quality of life and achieve sustainable growth, thanks to a team of more than 1,500 people committed to building a better world through technological research and innovation. That is why TECNALIA's research has a real impact on society and generates benefits in the form of quality of life and progress.

Its main scopes of action are: smart manufacturing, digital transformation, energy transition, sustainable mobility, health and food, urban ecosystem and circular economy.

In the latest awareness and positioning study carried out by the European Research Survey (ERS) in 2022, TECNALIA occupies the first position in R&D&i brand awareness.

3.2.2.2 Data spaces readiness

DENN has demonstrated significant readiness in adopting and implementing data spaces through its ongoing development and operationalization of DENNDATA. This platform integrates advanced data acquisition, storage, and processing capabilities, positioning DENN as a key contributor to the PLIADES project.

Experience with data integration and Big Data platforms:

DENNDATA was developed to centralize and analyse data from DENN's machines using proprietary tools, which enable real-time cloud integration and data transmission. The platform supports functionalities like production monitoring, energy consumption analysis, and vibration diagnostics, providing a robust foundation for integrating diverse data sources and enhancing interoperability.

Expertise in advanced analytics and AI integration:

DENNDATA incorporates AI-driven process parameter recommendations, which optimize manufacturing operations by analysing machine data and suggesting precise adjustments to improve efficiency and product quality. This capability demonstrates DENN's ability to implement AI within a data space framework effectively. Previous R&D initiatives provided DENN with expertise in developing predictive and diagnostic models, which enhance operational decision-making and data utilization.

Alignment with PLIADES objectives:

DENN has leveraged its expertise to adapt DENNDATA for complex multi-data space scenarios, aligning with the PLIADES project goals of integrating sustainability, operational, and manufacturing data lifecycles. The company's experience in deploying interoperable systems and managing industrial data spaces positions it as a capable partner in implementing scalable solutions that meet the project's technical and operational requirements.

Through these features, DENN demonstrates a high-level of readiness to contribute to the design and implementation of Data Spaces within the PLIADES framework, ensuring seamless integration, robust analytics, and actionable insights for industrial processes.

TECNALIA is member of various initiatives that comprise DSBA (Data Spaces Business Alliance). The research and implementation integration of Data Spaces has become of paramount importance for generating business impact through Research and Development activities. In addition, TECNALIA is member of various initiatives within DSSC (Data Spaces Support Centre). It also participates in various research projects that utilize data spaces, such as OASSES (Grant agreement ID: 101092702)[6], and DATAMITE (Grant agreement ID: 101092989)[7]. In terms of research activities, TECNALIA has authored various scientific papers related with data [1], [2], [3], [4].

3.2.2.3 Current business scenario

DENN has taken significant steps in optimizing its machine operations through the implementation of DENNDATA, to provide real-time insights and improve production efficiency. The platform collects a wide range of operational data from rotational forming machines, including production statistics, energy consumption, and machine diagnostics. This real-time information enables manufacturers to make informed decisions, automate production processes, and reduce machine downtime.

[6] https://oasees-project.eu/

[7] https://datamite-horizon.eu/

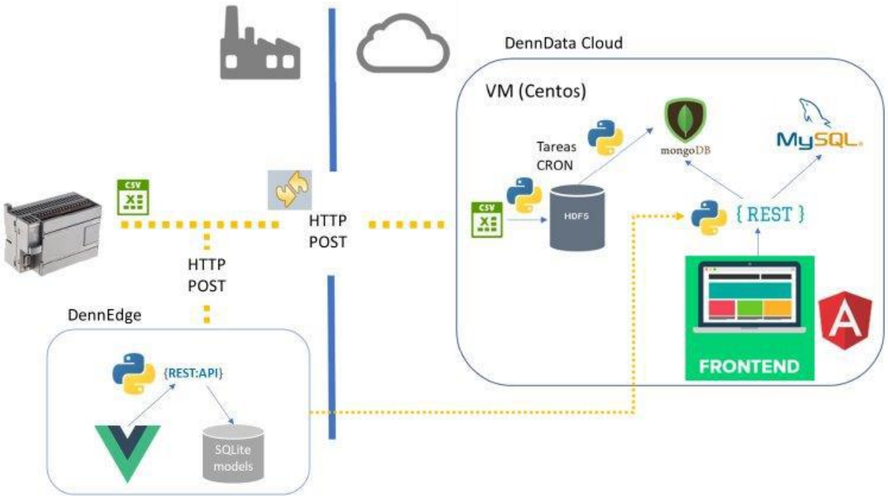

Figure 2 DENNDATA – Product digital platform's scheme

Through its R&D initiatives, DENN has incorporated additional sensors into its machines to enhance data acquisition capabilities. The development of proprietary tools for cloud-based data integration has enabled the creation of DENNDATA, a platform designed to facilitate seamless connectivity and analytical insights.

Core functionalities of DENNDATA include:

- Monitoring and analysing production metrics to evaluate machine output and efficiency.
- Tracking energy consumption at both macro and component levels.
- Vibration monitoring for assessing mechanical conditions and detecting potential anomalies.

DENNDATA integrates an AI-based process parameter recommender system, which provides precise adjustments for manufacturing settings, improving accuracy and efficiency, while reducing the need for manual intervention.

3.2.2.3.1 Increment DENN data platform penetration in reluctant sectors through data compliance and security

At present, DENNDATA is actively utilized for data collection and management across various sectors, including some automation applications. However, sectors such as aerospace and high-precision automation still show reluctance in adopting DENNDATA due to concerns over data security and privacy. The current process involves customers signing standard license agreements to access data services, which does not adequately address the flexibility required for data sharing or regulatory compliance in sensitive industries. The lack of integration with data spaces limits interoperability and scalability.

3.2.2.3.2 Data-driven lifecycle management for suppliers

Currently, lifecycle data from machine components like bearings and motors is used internally by DENN for performance optimization and predictive maintenance. Suppliers do not have direct access to this data, limiting their ability to make data-driven decisions about component durability, failure risks, or necessary design improvements. Maintenance schedules are typically determined by historical data or operator experience, which may not accurately reflect the real-time condition of the components.

3.2.2.3.3 Commercial expansion of DENNDATA for cross-industry data-driven optimization

DENNDATA is actively used within the manufacturing sector for operational optimization, predictive maintenance, and diagnostics. Some third-party services, such as data analysis and support, are offered to clients, but these services are limited to existing customers in the manufacturing industry. The platform's broader potential to serve other industries remains largely untapped.

3.2.2.4Pain points

The strategic objective of DENN is to include DENNDATA licenses in every commercialized machine increasing its revenues coming recurrently from services and acquiring more data from different sectors to be able to develop better AI services. Nevertheless, some relevant sectors where DENN participates are still reluctant to deploy DENNDATA due to data privacy and transparency issues, e.g., automation and aerospace. If DENN wants to reach its objectives, these concerns need compulsory to be dealt with. In this sense, technology developments leveraging on data spaces envisioned for UC1 constitute an extremely helpful tool.

On the other hand, besides increasing DENNDATA licensing in further sectors, another strategic objective for DENN is the increase of the value of DENNDATA platform through the promotion of AI-based services contained in the platform. Currently, penetration level of these services among DENNDATA users is still very low due to a set of barriers that hinder adoption. These barriers are associated with the effort required to labelled datasets required to train models, lack of automation in model management. DENN firmly believes that operationalization of data and model lifecycle management e.g., through active learning, data quality assessment, data and model drift detection or automation of model retraining triggering is one of the keys to foster adoption of current AI-based services and also to develop further ones.

3.2.2.5 Barriers

3.2.2.5.1 Legal framework

The industrial sector, especially in the context of integrating technologies such as AI models, predictive analytics, and automation, is subject to several important regulations that ensure safety, data protection, and environmental responsibility. These regulations carry significant legal implications for companies adopting advanced technologies, guiding both operational and compliance decisions. Additionally, to the regulations presented in 3.1.1, this use case is subject to:

- Machinery Regulation (EU 2023/1230): This regulation enforces strict safety requirements for machinery within the EU, particularly for companies integrating automation and AI. The regulation's legal implications ensure that machinery, including those driven by AI models, operates safely to protect workers. Compliance with this regulation is mandatory to prevent workplace accidents and ensure equipment reliability, and CE Marking is required to demonstrate adherence to these safety standards.

3.2.2.5.2 Social framework

The AI-Enabled Data Lifecycles Optimization trial introduces important social considerations, particularly in the context of privacy, security, and globalization within the industrial data management sector. These aspects directly impact how data is handled, shared, and secured through platforms like DENNDATA, and influence both the company's operations and broader societal expectations.

3.2.2.5.3 Economic framework

The current economic framework surrounding DENN's sector is strongly influenced by the integration of data-driven technologies and AI into industrial processes, particularly in the context of the AI-Enabled Data Lifecycles Optimization trial. As industrial data management becomes a critical aspect of optimizing manufacturing efficiency, DENNDATA plays a pivotal role in transforming how data is collected, stored, and monetized. This shift impacts various economic aspects of the company's operations.

One of the primary economic factors is the cost of implementing advanced data technologies like DENNDATA. The trial involves integrating AI and advanced data analytics into the company's infrastructure, which requires substantial investments in cybersecurity, infrastructure upgrades, and staff training. These upfront costs are necessary to ensure that the company remains competitive and can leverage the benefits of data-driven insights to improve operational efficiency. The cause of this economic pressure is the growing reliance on data for decision-making and the increased demand for secure, compliant data handling across industrial sectors.

Additionally, the potential to monetize industrial data through DENNDATA offers new revenue streams, particularly by selling data-driven insights to suppliers, customers, and partners. This data commercialization adds economic value but also introduces complexity in terms of ensuring compliance with regulations, such as GDPR and the EU Data Governance Act, which mandate secure and transparent data handling. The economic impact of complying with these regulations is significant, as companies must invest in the necessary tools and infrastructure to handle data responsibly while avoiding legal risks.

Finally, globalization and the increasing demand for interoperability between industrial systems across borders also affect the economic landscape. DENNDATA enables companies to collaborate internationally, but the challenges of navigating diverse data governance frameworks across regions, such as the EU DGA, require strategic investments in legal compliance and operational flexibility. This global push for data integration and secure data-sharing solutions drives further economic pressure on companies to innovate while maintaining compliance with varying regional standards.

3.2.3 Future scenario

3.2.3.1 Increment DENN data platform penetration in reluctant sectors through data compliance and security

Through PLIADES, the DENNDATA platform will evolve to include data-sharing capabilities that adhere to strict data compliance and governance standards, such as GDPR, ISO/IEC 27001, and the EU Data Governance Act. The integration of data spaces, will enable flexible data access and management through DS connectors, making it easier to establish revenue-sharing agreements with partners. This change aims to overcome barriers to adoption, allowing DENN to monetize its data insights in sectors previously hesitant to adopt the platform. DENNDATA will transition from a standard data service to a flexible, compliance-focused platform that can cater to multiple sectors, enhancing the commercial potential of the data and services offered. Key activities include:

- Data sharing with compliance assurance: Ensuring data governance standards are met while sharing information securely, complying with GDPR, ISO/IEC 27001, DGA, and other relevant standards.
- Revenue-sharing agreements: Establishing partnerships with suppliers, where DENN can benefit from the downstream use of insights derived from the shared data, including royalties for component improvements or new products developed using DENNDATA's analytics.

3.2.3.2 Data-driven lifecycle management for suppliers

The process will shift to a collaborative model where suppliers can securely access detailed lifecycle data of their components through DENNDATA. Data-driven insights such as vibration levels, temperature changes, and operational usage patterns will be shared with suppliers, enabling them to optimize maintenance schedules, reduce warranty claims, and enhance product design based on actual usage data. This transition will also create new revenue opportunities for DENN through data licensing or subscription-based access to enriched datasets. Moreover, partnerships will be established with suppliers to explore revenue-sharing models for the commercialization of data-driven insights. Key activities include:

- Component lifecycle analysis: Suppliers will be able to monitor the real-time condition and usage patterns of their components to improve product durability, with DENN potentially monetizing access to enriched data sets.
- Predictive maintenance coordination: Based on lifecycle data, maintenance schedules can be adjusted to better align with actual component wear and failure risks, enabling suppliers to optimize their services and potentially reduce warranty claims.
- Data sharing with compliance assurance: Ensuring data governance standards are met while sharing information securely, complying with GDPR, ISO/IEC 27001, DGA, and other relevant standards.
- Revenue-sharing agreements: Establishing partnerships with suppliers, where DENN can benefit from the downstream use of insights derived from the shared data, including royalties for component improvements or new products developed using DENNDATA's analytics.

These activities facilitate deeper integration of DENNDATA into the supply chain and enhance the sustainability of manufacturing operations through better component lifecycle management. By adopting a collaborative and monetization-focused approach, DENN can not only improve operational efficiency but also create new revenue streams through data-driven partnerships.

3.2.3.3Commercial expansion of DENNDATA for cross-industry data-driven optimization

As data spaces are integrated and the platform’s capabilities mature, DENNDATA will be positioned as a versatile data-driven optimization solution for various sectors, beyond its current manufacturing focus. The platform will offer tailored services and insights for industries such as logistics, energy management, and asset lifecycle optimization, using advanced AI models to support diverse applications. Commercialization will be achieved through subscription models, pay-per-use services, and/or collaborative data agreements with new industry partners. By facilitating secure data sharing and leveraging real-time analytics, DENNDATA will not only expand its market reach, but also improve interoperability, enabling companies across multiple sectors to harness data for operational efficiency. Key activities include:

- Operations optimization through data: Providing real-time insights to improve production efficiency and maintenance processes across multiple industries.
- Expansion of commercial services: Offering DENNDATA as a service platform to third parties, with customization possibilities for specific sectors.
- Compliance with data governance standards: Ensuring that data management practices remain aligned with international and EU regulations, such as the EU Data Governance Act and ISO/IEC 27001.

These activities aim to maximize data interoperability and open new business opportunities through the integration of industrial data spaces and other sectors.

3.2.3.4Future scenario data processes

Technology developments leveraging on data spaces envisioned for UC1 constitute an extremely helpful tool. Specifically, DENN seeks through the implementation of the UC to:

- Achieve full data compliance and security: Ensure that all data collected and processed via DENNDATA complies with GDPR, ISO/IEC 27001, and other relevant data governance frameworks, such as the EU Data Governance Act.
- Prepare for scalability and interoperability: Position DENNDATA to seamlessly integrate with broader industrial data spaces and platforms such as Gaia-X, enabling future scalability and cross-sector data sharing.

As described in the previous sections, the PLIADES framework will actively contribute to the business cases development, by implementing an improved data lifecycle management process that will boost the data value and usability for the different stakeholders.

Data filtering, considering here data harmonization performed in "Data generation and harmonization techniques for humans in mobility, industrial and healthcare data spaces", data quality assessment as part of "Design and automation of cross domain green data disposal process based on their quality or on legislation requirements", and "AI-Enabled data quality assessment and improvement strategies for heterogeneous data sources", and data drift evaluation according to "Human and data centric approaches to AI model development and maintenance".

Data annotation and human knowledge injection will be required as defined in "Empowering human-knowledge injection mechanisms in the data creation process" and "Human and data centric approaches to AI model development and maintenance".

Finally, **data sharing** will be implemented in the PLIADES framework, allowing the exchange of the generated data sets with different stakeholders. As a data provider, the sharing process will consist of incorporating the data sets into one or more data spaces, following the latest governance standards for a secure interchange. As a data consumer, data sharing will allow the access to the available datasets by agreeing to the specific policies for a correct use of the data.

3.2.4 Objectives

Table 2 UC 1 business objectives

Business objective	Description	Impact
Boost machine sales and increase DENNDATA integration	Incorporate DENNDATA into new machine models, increasing the proportion of machines sold with DENNDATA integration by addressing data security concerns and demonstrating compliance with data governance standards.	Results in a higher percentage of machines with DENNDATA, enhancing customer value, production efficiency, and data-driven decision-making.
Reduce critical maintenance time	Leverage predictive maintenance capabilities through DENNDATA to reduce the time machines remain offline during critical maintenance, enabling early detection and proactive scheduling.	Decreases downtime, improves machine uptime, and reduces maintenance costs, leading to higher operational efficiency.

Monetize data through strategic agreements with suppliers	Expand revenue streams by enabling data monetization through data-sharing agreements, allowing suppliers to optimize product designs and lifecycle management while DENN benefits from licensing or royalty-based models.	Generates new business revenue, establishes collaborative relationships with suppliers, and enhances manufacturing sustainability through data insights.
Enhance machine intelligence	Increase the level of "machine intelligence" by improving AI algorithms and data analytics, achieving higher levels of automation and decision-making accuracy with real-time data processing and human-in-the-loop approaches.	Enhances machine performance, optimizes resource usage, and reduces human intervention, leading to better production outcomes.
Expand market reach and improve interoperability	Extend DENNDATA's reach across multiple industries through data spaces, such as Gaia-X, enhancing interoperability and targeting sectors such as logistics, energy, and other data-intensive industries.	Increases scalability, enables entry into new sectors, and expands the customer base, driving long-term growth.
Drive innovation and ensure data compliance	Maintain compliance with data governance standards (GDPR, ISO/IEC 27001, EU Data Governance Act), while facilitating secure data-sharing across stakeholders to foster innovation and create new business opportunities.	Supports sustainable growth, promotes trust in data-driven processes, and enables the exploration of new technological applications.

3.2.5 Actors involved

Table 3 UC1 actors involved

Actor	Business area	Type of impact	Description of the impact
Machine operators	Manufacturing	Direct	Operators will benefit from automated machine adjustments and predictive maintenance alerts, reducing manual interventions.
Maintenance technicians	Technical support	Direct	Maintenance staff will have access to real-time machine data for predictive maintenance, allowing for quicker issue resolution and less downtime.
Production managers	Management	Direct	Managers will gain better oversight of production efficiency and machine performance, aiding in decision-making and process optimization.

Data analysts	Others/ Data Analysis	Indirect	Analysts will use the data from DENNDATA to improve machine learning models, optimize production workflows, and predict potential failures.
Software developers	IT department	Direct	Software developers will have access to advanced tools.
Suppliers	Purchasing	Indirect	Suppliers will have access to lifecycle data of components, allowing them to optimize maintenance schedules and improve product quality.
Customers	Customer relationships	Indirect	Customers will benefit from improved product quality and reduced machine downtime, leading to increased satisfaction and potentially longer contracts.

3.2.6 User & system requirements

Table 4 UC1 requirements

Business objective	Requirement
Enhance operational efficiency and reduce costs	Enable real-time monitoring of machine performance to detect inefficiencies
	Automate data collection for predictive maintenance scheduling
	Optimize energy consumption tracking through machine data analytics
	Integrate automated alerts for abnormal machine behaviours
	Provide real-time updates on inventory levels to optimize stock management
	Implement data-driven scheduling for maintenance based on actual usage patterns
	Establish a feedback loop for continuous machine performance improvements
	Allow dynamic adjustment of production parameters to reduce waste
	Monitor operational KPIs automatically to identify areas for cost reduction
	Provide role-specific dashboards to improve decision-making processes
Improve product quality and flexibility	Implement automated quality checks based on real-time machine data
	Enable flexible machine settings adjustments based on varying production requirements

	Provide adaptive control for material specifications based on data-driven insights
	Use historical data to anticipate production bottlenecks and adjust processes proactively
	Integrate quality data from previous production runs to improve current batch consistency
	Facilitate collaborative adjustments between operators and quality managers for improved outcomes
	Use predictive algorithms to adjust machine parameters before defects occur
	Enhance data visualization tools for real-time monitoring of quality metrics
	Create alerts for deviations in material quality based on set thresholds
	Adjust production schedules dynamically to account for material and product specification changes
Drive innovation and ensure data compliance	Establish a secure data governance framework aligned with GDPR and ISO/IEC 27001
	Enable secure data sharing protocols across stakeholders, maintaining data sovereignty
	Integrate human-in-the-loop methodologies to refine AI algorithms
	Create data anonymization techniques for sharing insights without compromising sensitive information
	Automate compliance checks for data access requests
	Develop tools for tracking data usage and audit trails for accountability
	Use advanced encryption techniques for data stored on cloud servers
	Establish a data quality monitoring framework to ensure continuous improvement
	Provide users with real-time compliance notifications for data operations
	Enable data-driven insights for regulatory reporting
Increase revenue streams and business opportunities	Establish data-sharing agreements with key stakeholders
	Develop pricing models for data access and premium features
	Enable dynamic billing systems based on data usage
	Integrate data analytics to identify new revenue streams

	Develop marketing strategies for new services derived from DENNDATA
	Establish partner programs for shared monetization of generated data
	Implement dashboards for tracking monetization KPIs
	Use AI-driven insights to create personalized service offerings for clients
	Automate royalty payments for partners using data monetization
	Establish frameworks for licensing DENNDATA functionalities in new markets
Expand market reach and interoperability	Develop partnerships with technology companies to integrate DENNDATA into their solutions
	Identify potential new sectors for DENNDATA expansion, such as healthcare or energy
	Implement pilot projects with early adopters in new markets
	Establish co-innovation programs with clients to develop new use cases for DENNDATA
	Develop training programs to help clients utilize DENNDATA for new business opportunities
	Use predictive analytics to identify emerging market trends
	Automate the generation of market reports based on data insights
	Leverage AI to explore cross-industry data opportunities
	Establish legal frameworks for co-development agreements with clients and partners
	Enable real-time collaboration tools for stakeholders working on new projects

3.2.7 Expected benefits and KPIs

The adoption of predictive maintenance and data-driven solutions, such as DENNDATA, presents significant benefits for the company and the industry. Increased uptime and improved operational efficiency of the rotational forming machines, achieved through real-time data insights, automated parameters' adjustments and predictive maintenance, lead to higher production efficiency, reduced downtime, and lower operational costs, driving significant financial benefits. These advancements result in a high impact on the industry by enhancing overall productivity and minimizing machine downtime., ensuring that equipment is serviced at optimal times. Additionally, the reduction in maintenance time and associated costs improves machine availability, extends equipment lifecycles, and reduces the financial burden on maintenance budgets, with high potential to optimize processes industry-wide. The integration of DENNDATA into machines—raising its adoption rate from 50% to 75%—enhances product quality, market presence, and customer satisfaction, encouraging the widespread adoption of data-driven solutions across traditional sectors. This contributes to business revenue growth by expanding sales and tapping into new market opportunities while enabling quick adaptation to changing production needs, setting a standard for revenue generation through data monetization and improving responsiveness in a rapidly evolving market landscape. Finally, it will improve sustainability in manufacturing, through the reduction of energy consumption and waste, aligning with zero-waste manufacturing initiatives and sustainability goals within Industry 4.0.

Table 5 UC1 Benefits and KPIs

Indicator	Description	Current value	Future expected value	Expected date of achievement
Increase in machines sold with DENNDATA	Increase the proportion of machines sold with integrated DENNDATA	50%	75%	Before 12 months after implementation
Reduction of critical maintenance time	Decrease in critical maintenance time (anticipating component replacement) to minimize downtime	5 months	1 month	Before 24 months after implementation
Data monetization with suppliers	Revenue generated through data-sharing agreements and strategic partnerships	0€	50k€	Before 18 months after implementation
Business revenue growth	Revenue growth achieved by expanding DENNDATA sales and entering new markets through data-driven value	Baseline revenue	+25% increase	Before 24 months after implementation

3.3 Use Case 2: Integrating data life cycles of service robot to improve HRI with end users

3.3.1 Use case overview

Assistive robots in healthcare settings can transform patient care by taking on tasks like patient transfer, receptionist duties, and patient monitoring. Such robots can help reduce physical strain on healthcare workers, ensuring safer, more efficient patient handling, particularly for those with mobility challenges. For instance, robots can assist in lifting or transferring patients, minimizing the risk of injury to both staff and patients. As healthcare systems face growing demand due to an aging population, especially with the rise in elderly patients needing regular care, robots offer a scalable solution to address staffing shortages and improve the overall quality of care. Additionally, robots equipped with monitoring systems can track patient data more efficiently than the human eye, enhancing early intervention and patient safety. This use case focuses on improving the interaction between robots and patients. Assistive service robots, either for rehabilitation and patient support purposes in hospitals or for eldercare support in nursery or domestic settings, need data, collected during their actual operation, for their Human-Robot Interaction (HRI) capabilities to be trained and improved over time. The individual specificities of each human make each HRI session rather unique while the recent advances in Deep Learning (DL) methods highlight the need for large volumes of data to train AI systems that operate with highly advanced capabilities of personalization and effectiveness during the interaction.

This use case explores the collection of massive data that can be used to train human activity, behaviour monitoring and efficient HRI methods for: i) patient support and rehabilitation robots and ii) social assistive service robots

The data providers as well as data users will be BOR and CERTH within the healthcare Data Space. I4BD will assist, by providing additional equipment for the robots and a simulation testbed for data collection.

3.3.2 Background

3.3.2.1 Participants

Blue Ocean Robotics (BOR) develops, produces and sells professional service robots primarily in healthcare. The company develops and introduces new innovative market-ready robots for healthcare, pharmaceutical, hospitality, and other global markets. The company develops robots from problem, idea and design to development, commercialization, and all the way to exit. Each robot brand is set up in its own subsidiary- venture company focused on commercialization, global distribution and growth. BOR will act as a data provider and data consumer within the Healthcare Data Space of PLIADES.

Centre for Research and Technology Hellas (**CERTH**) is a research organization working on various topics related to robotics and Human-Robot Interaction, including the development of human action recognition methods, human-aware navigation and assistive capabilities. The healthcare domain is a setting this research has targeted through the years, and PLIADES will be an opportunity to further expand and improve the developed methods, while also producing valuable data for use in data spaces. CERTH will also act as a data provider and data consumer within the Healthcare Data Space of PLIADES.

I4byDesign (I4BD) is a Competence Centre focused on helping SMEs towards the adoption of Industry 4.0 & Logistics 4.0 technologies. Its vision is to combine know-how, experience, and technical expertise of its shareholders on Industry 4.0 technologies, to promote the application of integrated technological solutions. Its mission is to contribute, in National and International level, to the acceleration of digital and technological transformation of the Greek industrial ecosystem, focused on SMEs. In this use case, I4BD will assist by providing additional equipment for the robots and a simulation testbed for data collection.

3.3.2.2 Data spaces readiness

BOR has not worked with Data Spaces before joining the PLIADES project and thus sees the project as business opportunity to improve its robots with the use of Data Spaces.

CERTH has no prior experience on the usage of Data Spaces, but in contrast has great experience in the development of AI methods for mobility applications.

I4BD has no prior experience with Data Spaces. I4BD views Data Spaces as a strategic enabler for secure, interoperable data sharing, aligning with its mission to deliver cutting-edge solutions. By participating in PLIADES, I4BD aims to enhance the expertise of its personnel, build strategic collaborations with key players from around EU and expand its business opportunities by providing competitive and innovative solutions.

3.3.2.3Current business scenario

Two business scenarios have been identified within Use Case 2 – "HRI in rehabilitation and HRI with social assistive robots". The first one refers to the improvement of human-robot interaction when robotic systems are used for patient rehabilitation. The second scenario relates to the enhancement of human-robot interaction in a healthcare environment, where robots are used to monitor or guide patients.

3.3.2.3.1 HRI in rehabilitation

The current process in which the use case is going to be implemented is a rehabilitation scenario, where the robot is assisting a patient to perform a rehabilitation exercise. Currently, the robot is operated by a healthcare professional, as shown in Figure 3. It is envisioned that the robot can support the patient autonomously so that the healthcare professional can look after several patients at the same time. Currently, no data is collected by the robot and shared in a data space.

Figure 3 Current manual process gait training with PTR robot

3.3.2.3.2 HRI in patient monitoring

Socially assistive robots can be valuable in supporting patients in a hospital setting. There are two processes in which the use case is going to be implemented: HRI in a receptionist scenario and HRI for patient monitoring. AI-based methods for accurately recognizing human actions, along with advanced methods for autonomous, human-aware navigation and interaction with the robot's user have been developed and tested by CERTH in various settings, with a focus on the healthcare domain during research projects. To that end, various sensors have been employed with occasional data collection, to train and improve the models.

Currently, on any CERTH robot that performs these tasks (see Figure 4), the data is processed and dropped during normal (live) operation without utilising it for later use. In cases where the captured data should be stored or exploited, a process is explicitly triggered for that purpose. Evidently, this is not yet an established process but could become a realistic service in the future with the help of PLIADES technologies.

Figure 4 Current robots from CERTH performing human action recognition & receptionist tasks

3.3.2.3.3 Current data processes

The data processes that are common across the two scenarios are the following:

- Data collection
- Data filtering
- Data synchronization
- Data annotation
- Data sharing

Data collection refers to the production and storage of raw multimodal sensor streams from autonomous robots and their input devices. This process starts on demand and stores the data locally or streams the sensor measurements and completes with the storage of these data into a server. This currently leads to mass amount of data, making it challenging to select the most useful parts of the collected data with the optimal quality that will assist the AI development.

During the data creation step, the format of the generated data may not the same among the various robots and partners. Even though BOR and CERTH utilize ROS software stacks, the created datasets might not be created in the same way, thus leading to incompatibilities.

These data are saved in local servers of the data creator without considering any sustainable practices, leading to massive incoming data and potential burden when shared.

Data filtering refers to the execution of processes that improve the data quality by removing duplicates, excluding low-quality or noisy frames (e.g., blurred images, silent chunks of audio), removing corrupted frames and generally eliminating irrelevant information that could lead to potential confusion in the system. Currently, data filtering is performed offline as part of the data pre-processing steps. This process involves manual reviewing of the collected data by human experts, followed by manual data exclusion in cases that it is needed.

Data synchronization refers to the process of establishing consistency between the data originating from different sensors on the same agent. Robots equipped with multiple sensors are heavily reliant on the accurate synchronization of the sensor streams, so that all of them can be used for processing with accurate measurements. Existing tools, such as message filters provided for ROS-based infrastructures provide such functionalities. Since robots from BOR and CERTH use ROS, the only issue will be to ensure that the related stamps are always saved along with the data.

Data annotation refers to the process of generating labels for the collected data. These labels might refer to various AI-based tasks, such as human recognition and activity monitoring, or might be the recognition of events. The format of the labels can be either visual (image masks), text description or even audio labels. Also, the annotation can be related to different types of data and can have different formats for each one, e.g., the labels for speech data are different that the labels for image data.

Currently, the data annotation process is mostly performed manually using some publicly available tools, either online or offline to extract the corresponding labels, or using some semi-automatic AI-based labelling pipelines that are executed in the cloud; however, they always need human verification.

Another aspect of the data annotation process is the format of the labels. Currently, there is no standard followed and there are several available formats that are used by AI-based methods, such as YOLO format, COCO format etc.

Data sharing refers to the data interchange among various partners who share common needs and would benefit from having access to data from different conditions, sensors or robots.

Currently, data sharing is not an automated process and also requires great effort to adapt the available datasets, so that they will be usable. From the perspective of data providers, existing data sharing practice is to upload the generated datasets on a publicly available server, so that everyone has access to the data or to private servers, where access is selectively provided. In both approaches, this process requires direct contact between the two parties (provider and consumer).

On the side of data consumers, the data are downloaded locally to a computing unit, their integrity cannot be guaranteed, while improvement by the data provider cannot be performed online but requires complete transfer of the dataset. Also, the data consumer needs to manually search for the available datasets, through online catalogs and select the optimal dataset, based on samples provided and their description.

3.3.2.4Pain Points

Social and ethical aspects: Ensuring the safety of individuals passing near the robot during its trials is a critical concern. This raises ethical questions about how to protect people while testing the robot in real-world environments. The robot must also be designed to operate in a way that respects societal norms and privacy, especially during data collection activities.

Technical aspects: The robot must be capable of determining whether a situation is challenging enough to require external assistance. Capturing data dynamically while ensuring compliance with privacy-preserving practices and GDPR regulations is another significant challenge. Identifying and recognizing the trigger points that signal the need to begin automatic data collection and storage is essential. Additionally, processes and communication channels must be established for transferring data from the robot to designated Data Spaces. Managing the substantial amount of data that needs to be stored presents a further technical challenge. If data is stored locally, disk space could become an issue, while streaming it externally may lead to bandwidth and networking constraints.

3.3.2.5Barriers

3.3.2.5.1 Legal framework

The EU’s legal framework for healthcare robotics and AI combines medical device regulations, data protection laws, liability standards, and ethical guidelines. These regulations ensure that robotic systems and AI applications are safe, effective, transparent, and accountable, while also safeguarding patient rights and privacy. Additionally, to the regulations presented in 3.1.1, this use case is subject to:

- Medical Device Regulation (MDR): In the EU, robotics and AI used in healthcare often fall under medical device regulations. The MDR (EU 2017/745) ensures that medical devices, including robotic systems and AI applications, are safe and effective. These regulations require conformity assessment before market entry and ongoing post-market surveillance.
- Liability and Accountability: Liability for robotic and AI systems in healthcare involves on the one hand product liability: Manufacturers are responsible for defects that harm patients. On the other hand, it includes professional liability: healthcare providers remain accountable for ensuring the safe use of AI and robotics, requiring human oversight of clinical decisions.
- CE Marking and Conformity Assessment: For market access in the EU, healthcare robots and AI systems must receive CE marking, demonstrating compliance with safety and performance standards. High-risk devices require approval through a Notified Body. As a standard, ISO/IEC 13482:2014 is important, which also focuses on the safety requirements for personal care robots.

3.3.2.5.2 Social framework

The increasing use of robots in therapeutic settings presents various complex and unresolved legal and ethical challenges. These include potential threats to patients' autonomy, human dignity, and trust in healthcare, as well as the risks associated with inaccurate or malfunctioning technology, which could have life-threatening consequences. Additionally, the extensive reliance on personal and sensitive health data raises significant concerns about privacy, while the cyber-physical nature of these robots introduces new vulnerabilities related to data security. Furthermore, there are open questions about how to obtain truly informed consent when medical decisions are guided by AI systems whose decision-making processes may be difficult for both patients and healthcare providers to fully understand or assess. As these technologies evolve, the need for comprehensive ethical frameworks and robust legal safeguards becomes increasingly urgent to protect patients' rights and well-being. Despite the legal frameworks outlined in the previous section, there are aspects still not covered by established standards e.g., concerning the communication between robots/AI and patients. Current standards are primarily concerned with the physical safety of medical devices. However, standards currently miss to cover the specific dangers to mental health and other patient rights arising from e.g., HRI.

3.3.2.5.3 Economic framework

The integration of robots into healthcare settings, particularly hospitals and retirement homes, presents a unique set of challenges due to the complexity of the technology. While robots have the potential to significantly improve efficiency and patient care, their initial implementation is far from straightforward. Beyond the considerable financial investment required to acquire and set up robotic systems, these technologies demand substantial time and attention from healthcare professionals, particularly during the learning and adaptation phases. Training staff to operate and maintain the robots, integrating them into existing workflows, and ensuring they complement human care providers all require significant time and effort.

This challenge is further increased by demographic shifts in the healthcare workforce. With an ageing population and growing demand for healthcare services, the availability of skilled healthcare workers is already strained. As a result, the time available for staff to invest in learning new technologies, including robots, is increasingly limited. This growing pressure on healthcare resources makes the initial deployment of robots more difficult, as it competes with the urgent need to address immediate patient care demands.

However, the PLIADES technology can support the long-term economic benefits of robotic implementation. By optimizing the deployment process and reducing the time needed for staff to adapt, this innovation will help alleviate some of the initial burdens. Once the robot is fully integrated into the healthcare setting, it will allow healthcare professionals to attend to more patients and increase the quality of care through a smooth support by the robot, ultimately contributing to a more efficient, cost-effective, and sustainable healthcare system.

3.3.3 Future scenario

3.3.3.1 HRI in rehabilitation

The fundamental process remains the same as before, but the key difference lies in the robot's role in supporting the patient, freeing the physiotherapist to focus more directly on the patient's progress. In this future use case, enabled by the PLIADES project, the process unfolds as follows:

First, the robot navigates to the designated gait-training area, such as a long corridor or a spacious rehabilitation room. Upon arrival, staff members assist the patient in securely attaching to the harness connected to the robot. The patient should train walking in a secure manner, but needs to be able to walk safely on their own.

Next, the robot's graphical user interface (GUI) offers various operational modes, including autonomous movement at a pre-set speed, facilitating the gait training. The staff selects the mode that best suits the individual patient's needs. Once the chosen mode is activated, the robot begins moving, allowing the staff to concentrate fully on the patient's performance and well-being, rather than manually controlling the robot with a joystick. Throughout the training session, gait data is continuously collected.

The PLIADES technologies also enable several additional interaction possibilities to enhance the training experience. One such possibility involves the robot projecting patterns onto the floor or utilizing strategically placed stickers to create a more challenging gait training environment. In a more advanced iteration, the robot could even recognize the patient's success in navigating these challenges and provide feedback, either visually through lights or audibly through spoken announcements.

Furthermore, the robot can interact directly with the patient through voice prompts, such as inquiring about their comfort level or the pace of the robot. The staff can pre-select or adjust the interaction mode (e.g., a proactive robot offering frequent interaction or a more passive mode with minimal interaction) either before or during the training session.

Finally, the gait data gathered during the training is analyzed. By comparing this data with data from other patients and their respective treatments, the robot, utilizing AI, can generate a suggested plan for the patient's next rehabilitation training session.

3.3.3.2 HRI in patient monitoring

Receptionist scenario: When a patient arrives at the reception hall, they interact with a receptionist robot, communicating either through voice commands or via a graphical user interface (GUI). Depending on the type of interaction, the robot will either provide the requested information of guide the patient. More specifically:

a. The robot begins an interactive session with the patient to provide information based on the patient's requests. This might involve answering questions about directions, services, or other relevant information.
b. The robot guides the patient to a specific room or part of the building. During this escort, the robot dynamically adjusts its speed to keep pace with the patient or pauses to wait for them if necessary. This adaptive movement is a key feature, ensuring the patient feels comfortable and supported. The system also allows the user to start and stop the robot using basic gestures, providing additional control. Furthermore, the robot is designed to recognize the patient even in crowded operational spaces using gait recognition technology.

c. Finally, upon completion of the task—either by a prompt from the user or upon reaching the destination—the robot returns to its initial state and position in the reception hall, awaiting the next patient interaction.

Patient monitoring scenario: The patient monitoring process begins when a clinician or caregiver assigns a monitoring task to the robot through its GUI. The caregiver uses the GUI to provide essential information, including the patient's identity and current location.

Following task assignment, the robot autonomously navigates to the specified location, such as a room or part of the building. Upon arrival, the robot begins searching for the designated patient. This identification process can utilize various methods, including facial recognition and identification, static solutions like QR codes, or even verbal verification.

Once the patient is located, the robot initiates a monitoring period of a pre-determined duration, specified at the start of the task. During this time, the patient can use gestures to perform basic control functions of the robot or provide feedback.

Throughout the monitoring process, the robot analyzes the patient's activities and state, providing insights through a dedicated GUI. This interface allows caregivers to remotely observe relevant data and information.

Finally, upon completion of the monitoring period, or if the patient prompts the robot to stop, the robot returns to its initial state and position, awaiting the next task assignment.

In both cases, the technologies developed in PLIADES will facilitate the data process, enabling continuous and smart data collection from any modality of the robot that relates to the interaction with the user, such as GUI interactions, voice commands, gestures or gait speed. These in turn will be utilized to enhance any models used for HRI.

3.3.3.3Future scenario data processes

The processes that will be affected by the PLIADES framework are relevant to both business scenarios that belong to use case 2. The business processes are directly related to the data operations, which were mentioned in the previous section, and that are performed either by data providers or data consumers or both.

Data collection's evolution is related to the standardization of data formats that are used by the involved partners, as well as to the introduction of sustainable practices, leading to the green approaches for data collection.

The format of the collected data will be based on standardized formats used in robotics and HRI. Such formats are common file formats that are widely used for images, sound, videos, etc. such as .JPG, .MP4, .PNG, etc. Also, data collection will use a common format based on XML structures to describe the relative placement among different sensors and also any other necessary metadata that can assist the time synchronization of different data source.

Sustainable practices will be established in order to reduce the amount of data collected, but without compromising the information available. These AI-based methods will be executed either real-time, during the streaming of sensor data to a computing unit, either offline, as a post-processing step to the collected data.

Through the PLIADES framework, **data filtering** will undergo important improvements that will lead in improved performance of these processes, while also reducing the required time. Specifically, data filtering will be performed online, on the edge, by performing a sequence of filtering pipelines in a specified order. The outcome generated will be an updated version of the data with increased quality, along with quality metrics for various types of data.

These pipelines will evaluate the data quality and will automatically discard data that do not contribute in the collected data sets. AI-based approaches will be developed as part of the overall data processing pipeline and will take into account data quality aspects to perform automated data filtering. In addition, specific modules for privacy-related filtering will be developed, in order to either discard frames or edit them accordingly, in order to follow the privacy guidelines. Such methods include the blurring of faces in image data and other potential needs that will be extracted from the related tasks.

Regarding **data annotation**, the existing automated annotation methods will be further expanded with human knowledge injection methods. Also, annotation will be based on standardized formats. An appropriate ontology might also be used to match the data with real-world entities. Data annotation will also include the handling of metadata insertion, suitable for the improved discoverability of data. To this end, automated annotation tools will be developed.

Data sharing will undergo significant improvements by the incorporation of data spaces and the PLIADES framework. The previous data processes will prepare the data in a way that will allow seamless exchange among partners, allowing them to directly use them.

From the perspective of data providers, the data sharing process uses as input the data that have been generated by the previous processes of Collection, Filtering, Synchronization and Annotation. These data will be extended with semantics metadata that will contain all the essential information regarding the ownership of data, the access rights and the potential regulations of usage. These data with their metadata will be available to the AI-based brokering schemes and provided through the developed AI-based connectors. Data providers will also be able to actively maintain and update their provided datasets.

From the perspective of data consumers, they will first of all make use of the AI-based querying system and the AI-based brokers in order to identify all the available datasets within multiple dataspaces and receive an accurate suggestion on a dataset, according to their needs. Data consumers will also agree on the pre-defined policies for the data usage, as part of their connection on the dataspace, ensuring the proper use of data.

3.3.4 Objectives

Table 6 UC 2 business objectives

Business objective	Description	Impact
Enable seamless HRI by leveraging advanced Deep Learning methods to gather and analyse operational data from assistive service robots.	This will allow robots to continuously learn and adapt to the unique needs and behaviours of individual users. By e.g., enabling real-time data collection during robot operation, PLIADES will enhance the robots' ability to personalize interactions, improve rehabilitation outcomes, and optimize patient or elder care support over time.	Achieving this objective will improve the following: - Trust and acceptability of healthcare professional and patient in the robot - Economic benefits of robotic implementation, contributing to a more efficient, cost-effective, and sustainable healthcare system.
Ensure privacy, security, and traceability of data	With the introduction of data collection and data spaces, involved stakeholders will likely worry about privacy and security of	Achieving this objective will support the trust and

	the collected data, specifically if it is patient related. Thus, it will be necessary to implement robust measures to protect the confidentiality and integrity of the collected data, ensuring compliance with regulations and maintaining detailed records of data provenance and usage.	acceptability of the stakeholders.

3.3.5 Actors involved

Table 7 UC2 actors involved

Actor	Business area	Type of impact	Description of the impact
Healthcare provider - Management level	(Healthcare) Management	Direct	The management level at the healthcare provider will evaluate the technical results and assess (according to economic aspects, security, acceptability, etc.) whether it could be deployed at their facility in the future.
Healthcare professionals (physiotherapist, nurse, doctor)	Healthcare	Direct	Healthcare professionals will directly interact with the robot, potentially also oversee patient data and suggestions on possible healthcare programmes by the AI tools.
Patient	Healthcare	Direct	Patients will use and interact with the robot that is collecting the data and making use of the data from the data space.
Robot manufacturer	Manufacturing	Direct	The robot manufacturer is providing the data and using data from the data space to optimize the robot's ability to interact with humans.
Data integration and analysis specialists	IT	Direct	Data integration and analysis specialists will be responsible for integrating, analyzing, and preparing new data sources. They will benefit from improved data quality, accessibility, and more efficient research workflows, enabling them to conduct high-quality and accurate research.
IT security specialists from healthcare institution	IT	Direct	IT security specialists will be responsible for implementing and maintaining enhanced security and privacy measures. This will include managing encryption, access controls, and monitoring systems to ensure data integrity and compliance with regulations.
Research project managers	Management	Direct	Research project managers will oversee the integration of new data sources, the implementation of data sharing protocols, and

			coordination between departments and external partners. They ensure smooth data flows, adherence to timelines, and alignment with research objectives.
Compliance officers at healthcare institution	Legal	Direct	Compliance officers will ensure that all data handling and sharing practices comply with relevant regulations and standards. They will be involved in auditing data processes and maintaining compliance documentation, and ensuring adherence to privacy laws.
Financial administrators	Accounting	Indirect	Financial administrators will manage the financial aspects of the data integration and sharing initiatives. They will handle budgeting, funding allocation, and financial reporting related to these projects, ensuring proper financial management and accountability.

3.3.6 User & system requirements

Table 8 UC2 requirements

Business Objective	Requirement
Improved data quality and annotation consistency to enable seamless HRI by utilizing advanced ML/DL methods	Implement compression techniques for data transmission
	Optimize edge computing for on-board filtering processes
	Automated identification and discard of noisy and low-quality multimodal data
	Semi-automated methods for data quality estimation for multimodal data
	Automated metadata tagging during data collection
	Increased discoverability of data through context-aware metadata
	Automated analysis for cultural issues of human tracking data
Sustainable and cost-efficient data operations	Definition of policies for sustainable data operations
	Automated annotation processes for multimodal data
	Implement compression and filtering techniques to reduce data storage needs
	Automated identification and discard of redundant data
	Implement advanced encryption protocols for all data

Ensure privacy, security, and traceability of data	Automated monitoring of data adherence to privacy and quality standards
	Implement data anonymization techniques to protect sensitive information
	Set up a system for tracking data usage, sharing metrics, and maintaining version histories to ensure data traceability.

3.3.7 Expected benefits and KPIs

The expected benefits of UC2, focused on improving Human-Robot Interaction (HRI) between robots and patients, include enhanced patient care through autonomous and semi-autonomous robot operations in rehabilitation and monitoring scenarios, as well as optimized workforce efficiency by reducing the physical strain on healthcare workers, allowing them to focus on critical aspects of care. The integration of data spaces enables the collection and analysis of HRI-related data, driving continuous improvements in AI methods for human activity recognition and personalized interactions. Robots can be deployed across various settings, addressing staffing shortages, particularly in eldercare and hospitals, while greener data creation, filtering, and compression methods reduce energy consumption supporting sustainability goals. Additionally, the privacy-preserving and GDPR-compliant data storage and sharing processes foster innovation, enabling the development of effective interaction models and advancing healthcare robotics. These advancements not only improve scalability and efficiency but also align with broader goals of sustainability and innovation in healthcare.

Table 9 UC2 benefits and KPIs

Indicator	Description	Current value	Future expected value	Expected date of achievement
Increased usability	The robot will support healthcare professionals and patients during the operation, resulting in improved usability. The System Usability Score (SUS) is a well-known method to evaluate usability of a product.	69	80	Once technology is at TRL9
Increased efficiency of robot installation	With a more autonomous and interactive robot, installation at the healthcare site will be more efficient, especially if a robot is introduced to a new department internally or new personnel needs to become familiar with the robot. One aspect where the PTR Robot scored low on the SUS is “learn to use it quickly”. We expect this to increase after PLIADES.	3.5	4.5	Once technology is at TRL9

Increased trust and acceptability of healthcare professional and patient in the robot	The robots' ability to personalize interactions, improve rehabilitation outcomes, and optimize patient or eldercare support over time is expected to increase trust and acceptability of healthcare professionals and patients. One aspect where the PTR Robot scored low on the SUS was "confidence in operating the robot". We expect this to increase after PLIADES.	3	4	Once technology is at TRL9

3.4 Use Case 3: Integrating data lifecycles of personalized medicine services to improve diagnostic and prognostic clinical prediction models

3.4.1 Use case overview

The research and development of personalized medicine approaches requires the collection of a vast amount of personal healthcare data in a safe, anonymized and efficient manner. In this scope, MONDRAGON Corporation (with MU-EPS as cooperative university) together with CIC BioGUNE are currently developing a data platform capable of storing, managing and visualizing large volumes of data from over 11.000 participants. The collected data, which is longitudinal in nature, refer to biochemical, metabolomic, proteomic, genomic, lipidomic, lifestyle (nutritional and activity) and sociodemographic information while several Application Programming Interface (API) services have been developed for data consumption, research questions definition, data processing and visualization.

The goal of the use case is to provide advanced precision medicine services through the safe and secure of sharing data and AI models through a data space. In this project:

- Use data space connectors that attend to the International Data Spaces Association (IDSA) requirements through the context and cultural-aware data analysis methods for personalized medicine.
- Medicine data and AI models generated over those data will be then made available in the healthcare data space, via our AI-based data brokering schemes.
- The data users will be able to use our data integration framework to discover relevant data, further elaborate and use them appropriately to enhance their AI services that improve the diagnostic and prognostic clinical prediction models in the healthcare sector.

3.4.2 Background

3.4.2.1 Participants

Mondragon Unibertsitatea (**MU-EPS**) is a practical university immersed in business reality that combines training, research and knowledge transfer to improve the competitive position of companies and organizations. University that listens to companies and society and responds effectively and with quality to their challenges. It is part of MONDRAGON Corporation.

The **Center for Cooperative Research in Biosciences (CIC bioGUNE**), a member of the Basque Research & Technology Alliance (BRTA), located in the Science and Technology Park of Bizkaia, is a biomedical research organization conducting cutting-edge research at the interface of chemical, structural, molecular, and cellular biology. It focuses particularly on generating knowledge about the molecular bases of diseases for use in the development of new diagnostic and prognostic methods and advanced therapies.

3.4.2.2Data spaces readiness

MU-EPS has participated actively in research related to data spaces both in European-funded projects such as QU4LITY[8] and local/private funded projects such as the "Plan de Ciencia y Tecnologia" projects funded by Mondragon Corporation. Those projects where mainly oriented to the Industry4.0 domain and gave as a result the following publications [5], [6], [7] and [8].

3.4.2.3Current business scenario

All data are collected in a platform as depicted in Figure 5. In the *Data acquisition layer*, laboratories upload the files containing the results obtained from the analysis. Nextcloud is used to upload those files. Most of those results are obtained from two Bruker NMR (Nuclear Magnetic Resonance) spectrometers, 600 MHz AVANCE III HD and 600 MHz AVANCE NEO, both following In Vitro Diagnostic Regulation (IVDr). Those files are anonymized before entering the platform.

The current business process at CIC BioGUNE begins with the meticulous collection of data from a variety of laboratory analyses. These analyses include blood tests, urine tests, and other diagnostic procedures that generate valuable biochemical, metabolomic, proteomic, genomic, and sociodemographic information. The primary aim during this phase is to ensure that the data collected is comprehensive and accurate, providing a robust foundation for subsequent analysis. To safeguard the privacy of individuals and comply with relevant data protection regulations, all collected data undergoes a thorough anonymization process. This step is critical in stripping away any personally identifiable information, ensuring that the data cannot be traced back to any specific individual.

Once the data is anonymized, it is prepared for upload to a secure cloud platform. In the *Knowledge layer*, the information provided is extracted from the files and stored in data repositories (both structured and NoSQL databases). Java scripts are used for that purpose. Tools like Nextcloud are employed to facilitate this upload process. Nextcloud offers a reliable and secure environment for data storage, ensuring that the uploaded data is protected from unauthorized access and breaches. The centralized repository created in the cloud supports both structured and NoSQL databases, which are essential for managing and retrieving large volumes of diverse data efficiently. This setup allows for seamless data integration and accessibility, making it easier for researchers to work with the data.

Following the upload, the data enters the processing phase. This phase is crucial for transforming raw data into actionable insights. In the Data consumption layer, the platform provides a web application that enables the production of datasets by filtering information. The application queries databases according to the filters established by the user. The resulting datasets are visualized in the web application. Statistical calculations are performed on those datasets previous visualization to support researchers on interpreting those datasets. Additionally, those datasets can be downloaded to perform further analysis by means of the application of AI algorithms. Thus, the data is parsed and organized into appropriate repositories using sophisticated data processing algorithms and Java scripts. These algorithms perform initial data cleaning and transformation tasks, preparing the data for more detailed analysis. Statistical analysis techniques are then applied to the processed data to

[8] https://qu4lity-project.eu/

uncover patterns and trends. Additionally, Machine Learning (ML) techniques are employed to enhance the analysis, providing deeper insights and predictive capabilities.

Researchers and analysts can access the processed data through a user-friendly web application designed specifically for querying and visualization. This application plays a pivotal role in making the data accessible and interpretable. It allows users to filter the data based on specific criteria, enabling them to retrieve subsets of data relevant to their research questions. The application also provides advanced visualization tools that generate graphical representations of the data, such as charts and graphs. These visualizations help researchers to quickly grasp complex data patterns and draw meaningful conclusions. Furthermore, the application supports statistical calculations, offering researchers additional tools to validate their findings and derive insights.

In addition to online data analysis, the platform provides options for offline work. Researchers can download the datasets from the web application for more detailed offline analysis. This feature is particularly useful for conducting in-depth studies that require advanced computational resources or specialized software tools not available within the web application. By allowing data download, the platform ensures flexibility and supports various research methodologies.

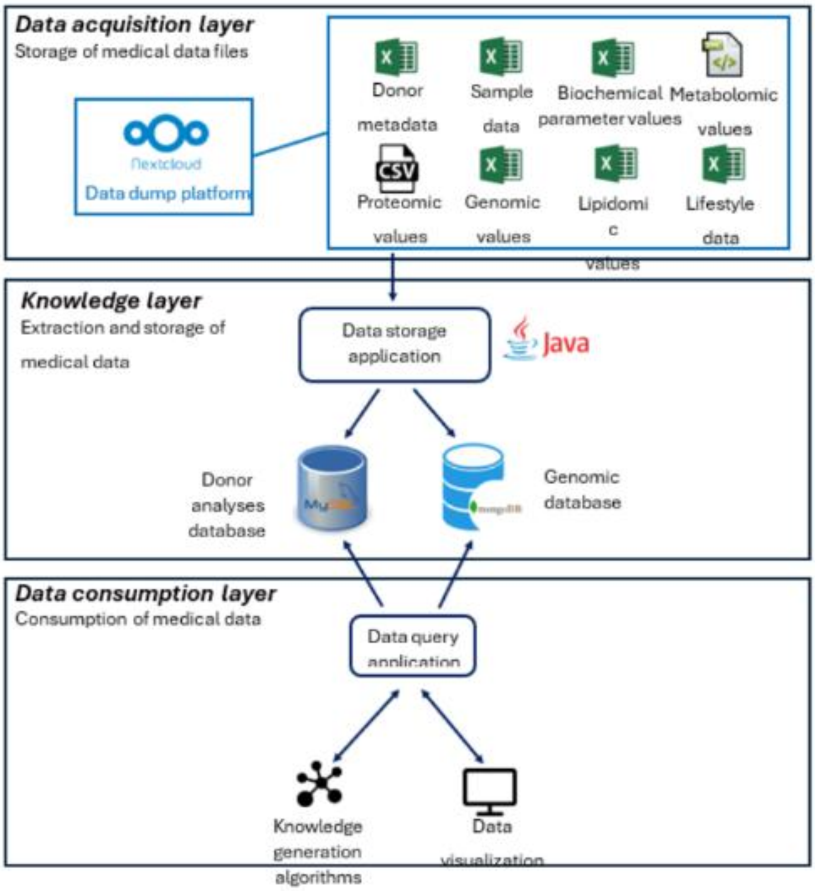

Figure 5 Current platform architecture of CIC BioGUNE

3.4.2.4Pain points

The challenges this architecture faces are the following:

- Define and implement shared data/model schemes. There are agreements among partners (labs-research centres) to share data but including a new one is a cumbersome task (new agreements).
- Enable the sharing of results and models so they can be used as inputs for new analyses.
- Safety and security issues for the implementation. In addition to the need of establishing agreements on the usage of data, security issues need to be improved.
- Enhance available data with third parties' data. Research would be improved with datasets from other parties (other laboratories or clinics) not enrolled in the program. These datasets are not always available as raw data and process data is offered. The metadata for these results is essential to provide meaningful information.

- Develop AI-based services for precision medicine. There is a need for further explore the capabilities of analysing these datasets by means of AI algorithms and extract knowledge that can be offered as services.
- Ensure traceability of data and previous results used in generating new outcomes or models. This requires maintaining detailed records of data sources, methodologies, and transformation processes to validate and replicate findings accurately.
- Correct versioning of data and results to accurately track changes and manage updates

3.4.2.5Barriers

3.4.2.5.1 Legal framework

The medical device sector is a strictly regulated sector. As later explained, this use case will provide an advanced service to safe share data an AI models. In this context, a medical device commercialization should comply with the following legislations/regulations, apart from the ones described in Section 3.1.1:

- Regulation (EU) 2017/745 (EU MDR): This regulation is designed to ensure higher levels of safety, efficacy, and transparency in the development of medical devices across the EU. The main objectives are: insurance the highest standards of safety and health protection; the enhancement of transparency and traceability of medical devices; and the alignment of EU regulations with technological advancements and international standards. Enforces strict safety requirements for medical devices within the EU, particularly for companies integrating automation and AI.

Medical solutions based on AI model should consider the following standards and certifications in the development phase:

- IEC 62304 - Software Life Cycle Processes: IEC 62304's goal is to ensure that software is safe, functional, and maintained throughout its life cycle, through appropriately documented architectural design and documented software validation process.
- ISO 13485: specifies the requirements for a Quality Management System (QMS) in the design, development, production, installation, and servicing of medical devices. It is focused on ensuring that medical devices meet both customer needs and regulatory requirements consistently.
- ISO 14971: Every medical device may inherently produce risks to the users. Therefore, it is crucial to show that all foreseeable risks have been identified, that controllable risks have been mitigated, and that systems are in place to detect and address potential risks.

3.4.2.5.2 Social framework

The social frameworks are described in Section 3.1.2.

3.4.2.5.3 Economic framework

The concept of precision medicine is rapidly growing in the medical community through a tremendous expansion of various advanced technologies. In terms of scientific production, in 2015, there were 1,737 articles with the term "Precision Medicine" in Pubmed, compared to just one article mentioning Precision Medicine in 2005.

The global market for precision medicine is estimated to reach $119 billion by 2026, with North America and Europe being the largest markets. In Europe, the precision medicine market generated a revenue of $20 billion in 2023, with a CAGR of 17.2% during the 2024-2030 forecast period.

Therapeutics was the largest revenue application segment; however, diagnostics is the most lucrative application segment registering the fastest growth during the forecast period. In terms of revenue, Europe region accounted for 23.3% of the global precision medicine market.

Precision medicine is mainly affected by predictive AI analytics, and, therefore, healthcare industry is no exception to the benefits offered by the latest advanced in AI and data sharing. It is helpful to reflect on the main fields benefiting from the expansion of predictive analytics, which will allow us to outline useful contingencies for future business scenarios:

- Operational efficiency in decision-making: Predictive analytics enables the restructuring of Business Intelligence strategies, facilitating access to large volumes of data. Real-time reporting is relatively new, but it is gradually being integrated into the healthcare industry, such as defining new treatment strategies or better resources allocation and planning.
- Accuracy in diagnosis and treatment in primary care: AI-based analytics allows the evaluation of forecasts and data so that healthcare professionals can find answers to incurable diseases. As a result, global mortality rates could gradually decrease.
- Greater knowledge to improve treatment for at-risk groups: The growing digitization of electronic health records, along with legal requirements for performance reporting, provides incredibly valuable datasets to gain insights into the health of specific social groups.

For the development of these scenarios, large volume of data is required and, therefore, secure data sharing strategies are required as those proposed in this project.

3.4.3 Future scenario

The current process, while robust, is geared towards ensuring data integrity, security, and accessibility. It sets the foundation for collaborative research efforts, enabling researchers to utilize high-quality data for their studies. However, as the field of data analysis and biomedical research evolves, there are opportunities to enhance this process further. The integration of new data sources, improved security measures, and the development of advanced AI-based services represent the next steps in this evolution, promising to elevate the quality and impact of the research conducted at CICbioGUNE.

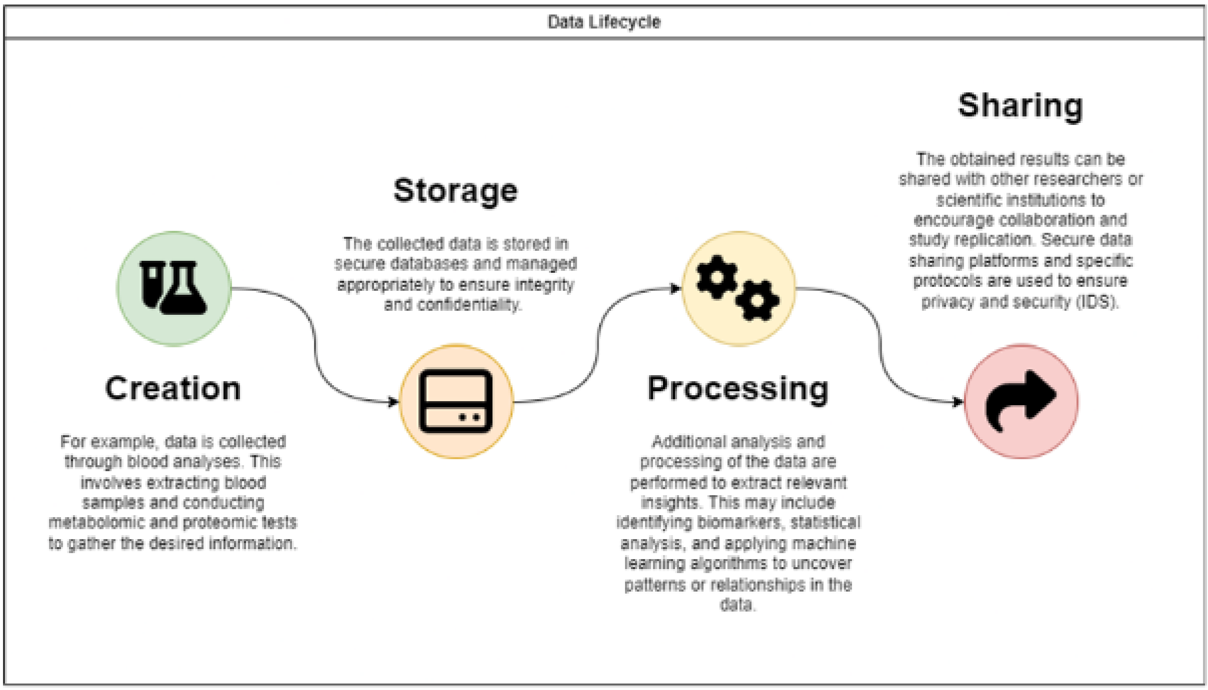

Figure 6 Data lifecycle

Two business scenarios are envisioned for the future. Before analysing each scenario, we present the general scenario data lifecycle (Figure 6) and the overall architecture (Figure 7). Four stages are visioned for the data lifecycle. Data is collected by means of analyses (blood, urine, etc.) or surveys in the Creation stage. Machines might be involved in extracting information from those tests. Data is stored in secure databases and managed appropriately to ensure integrity and confidentiality (Storage stage). Relevant insights are extracted from those data by means of statistical or machine learning techniques (Processing stage). Results (models or data) can be shared with other institutions using data space technology to ensure privacy, traceability and security.

The architecture (Figure 7) considers that each laboratory, research center or clinic is connected through data spaces. Each institution has a data space connector and can share data through the data space. Contracts and agreements on data sharing are established among partners according to the protocols and technologies offered by the data space, which is also responsible for providing identity, logging transactions and offering mediation.

Analysis of data can be performed by any organization or can be delegated to those organizations with the capability of conducting these tasks. Considering that data in this context is highly sensible, models could be shared among participants instead of data. AI implementation at each partner own premises and federated learning acquires high relevance.

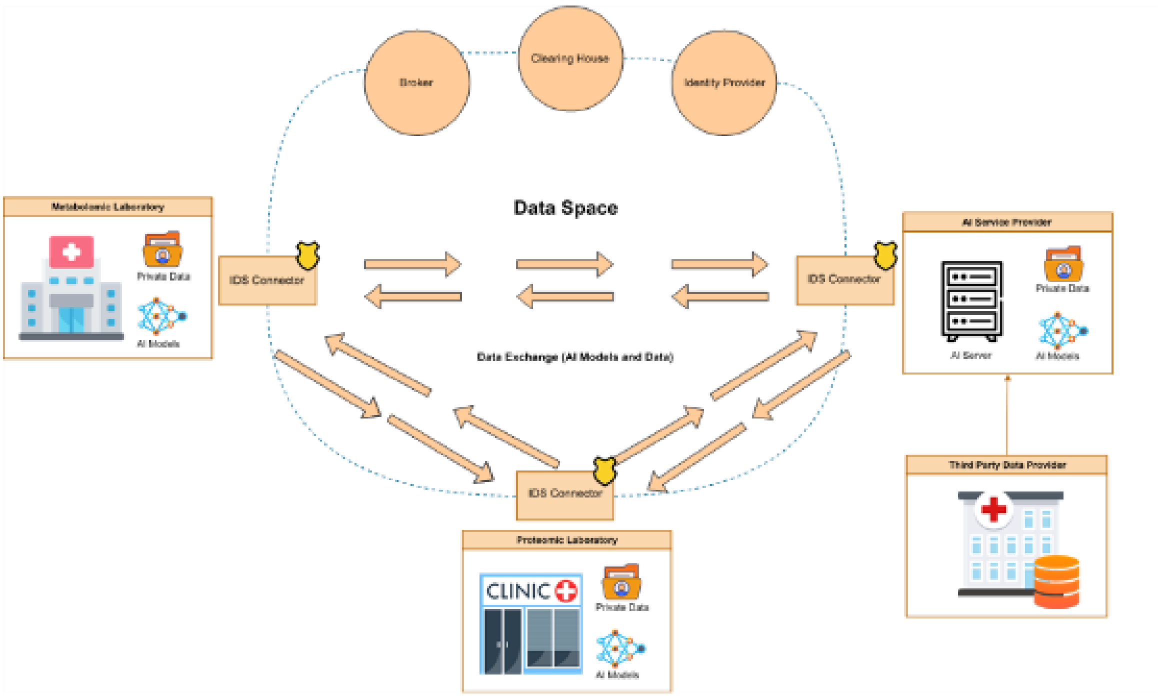

Figure 7 Data space-based architecture

3.4.3.1 Sharing data and models among partners

In this scenario, two or more entities collaborate and share data for mutual benefit, which may include data, outcomes, or even machine learning models. Data sharing is provided through the data space, as presented in the overall architecture.

As shown in Figure 8 models can be calculated and shared among institutions. The sharing of models enables the implementation of federated learning.

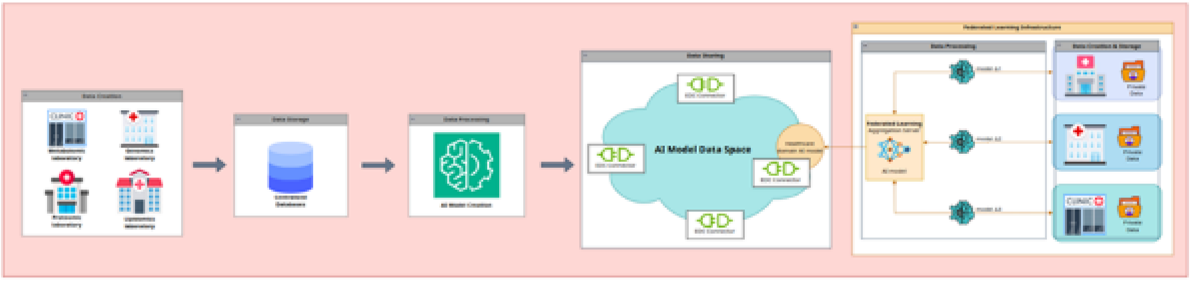

Figure 8 Sharing data and models scenario

3.4.3.2 Collaborative data/service exchange and third-party data acquisition

This scenario considers the exploitation of the precision medicine results obtained from the data with other institutions or clinics. This scenario involves sharing data with partners through APIs (services) and acquiring data from third-party companies for enrichment or research purposes.

For the first case, the results will be offered as services through data spaces as shown in the upper part of Figure 9. From the data, available models will be calculated. Those models will be deployed in a server in such a way that will provide answers to request in a client-server manner. These models will give recommendations or results to requests where the client provides input data. These input data will be shared through data spaces, assuring privacy, security and control usage. The results given by the server will be also shared through the data spaces.

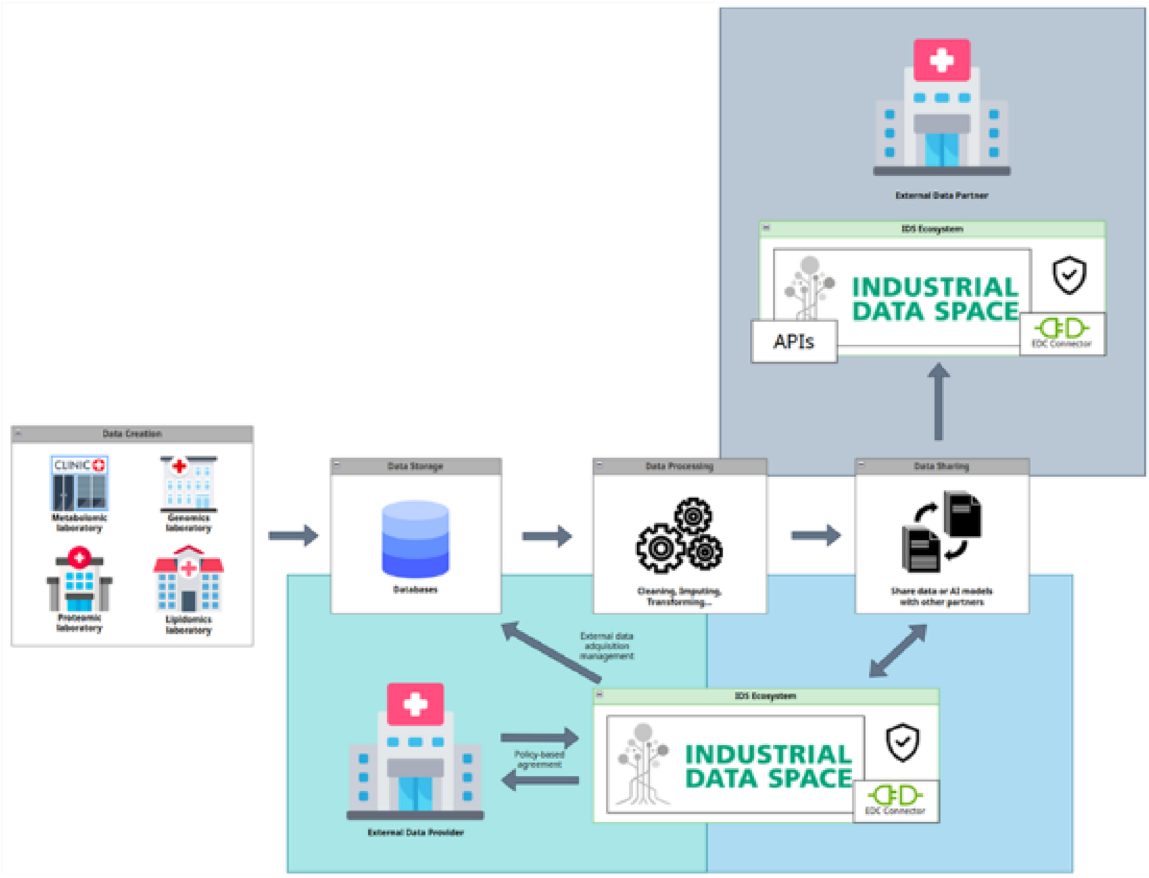

Figure 9 Collaborative data/service exchange and third-party data acquisition scenario

For the second case, third party datasets will be collected through data spaces, as shown in the lower part of Figure 9. These datasets will contribute to improving the models generated at platform level by means of federated learning.

3.4.3.3 Future scenario data processes

The expected evolution of the data processes involves several significant enhancements aimed at integrating new data sources, improving security measures, and developing advanced AI-based services.

In the future, the process will incorporate data from third-party sources to enrich the existing datasets. This integration will involve establishing new collaborations with additional institutions, thereby expanding the scope and depth of data available for analysis. These new data sources could include other research laboratories, clinical data from hospitals, or even data from wearable health devices. The inclusion of diverse data types will provide a more comprehensive view and enable more robust analyses. To facilitate this, streamlined agreements and protocols will be developed to ensure that new data can be integrated efficiently and without extensive bureaucratic delays. This approach will enhance the overall quality and comprehensiveness of the research data, allowing for more detailed and multi-faceted studies.

To address increasing concerns about data security and privacy, the future process will implement advanced security measures. These measures will include the use of sophisticated encryption techniques to protect data both at rest and in transit. Secure access controls will ensure that only authorized personnel can access sensitive data, and continuous monitoring will be put in place to detect and respond to potential security breaches in real-time. By strengthening the security framework, the process will better protect sensitive information and maintain the trust of data providers and users. Ensuring compliance with international data protection regulations, such as GDPR will also be a priority, with regular audits and updates to security protocols.

The evolution of the business process will also see the development of new AI-based services designed to further process and analyse the collected data. Advanced machine learning algorithms will be employed to identify patterns, make predictions, and generate insights that were previously unattainable. These AI services will not only enhance the analytical capabilities of the research teams but also automate many aspects of data processing and analysis, increasing efficiency and reducing the potential for human error. The results generated from these AI analyses will be shared through APIs, allowing other institutions to access and utilize the enhanced models for their research purposes. This will enable a more collaborative research environment where institutions can build on each other's findings. Additionally, the implementation of federated learning will be considered, allowing models to be trained across multiple datasets without the need for data to leave its original location, thus preserving privacy while still benefiting from a larger dataset.

A key component of the future process is the introduction of an advanced data space infrastructure to facilitate secure data sharing and collaboration. This infrastructure will provide a standardized framework for data exchange, ensuring that data sharing activities comply with established agreements on data usage among partners. The data space will support interoperability between different data systems and platforms, making it easier for institutions to collaborate without technical barriers. It will include mechanisms for identity management, transaction logging, and mediation services to ensure that data sharing is secure, transparent, and efficient. By adopting this comprehensive approach, the evolved business process will significantly enhance the ability to conduct high-quality, collaborative research, fostering innovation and improving the overall impact of the research efforts at CIC BioGUNE.

The enhancement of the precision medicine platform currently used will be achieved by implementing the functionality of some of the components identified in **Deliverable D2.3**. The specific actions and components to implement in the platform are presented next indicating in which stage of the process they will be included:

- Data collection: The scripts used for data collection in the actual version of the platform will be incorporated to the "**Data Harmonization and Pipeline Orchestration**" component described in D2.3. This will enable a better structure of the data process and include access to additional components. It will also automatize the data collection process.
- Data filtering: Efforts will be made to include the Web Application tool used to produce Data Products (datasets) and the catalogue into the previous component (Data pipeline).
- Data annotation: Datasets or Data Products and AI models will be labelled using the "**Active Learning**" component functionality. This component addresses the need for annotation and the production of metadata through human knowledge injection. This will enhance the Web Application tool enabling the addition of metadata and annotations over datasets and AI models. This component also complies with the "**Proven Ontology Manager**" that proposes the use of semantic technologies for labelling.
- Data sharing: Data sharing among the use case stakeholders will be implemented using the data space infrastructure proposed in PLIADES. In special, those related to data space connectors such as "**Interoperability Sharing Strategies/Protocol**".
- Data exploitation: For data exploitation the use case proposes to use AI techniques (Transfer learning, Federated learning …). The components identified in deliverable D2.3 for this purpose are related to WP6. The use case will closely follow the recommendations and components created in this work package. Specially the following ones: "**Explainable Model Creation**", "**Federated Learning**" and "**Data Aggregation, Analytics and Visualization (DAAV)**". This final component is related to the Web Application Tool available in the platform since it will add statistical data to the solution.

3.4.4 Objectives

Table 10 UC 3 business objectives

Business objective	Description	Impact
Facilitate integration and sharing of health research data and results	Streamline the processes for integrating diverse health data sources and enable seamless sharing of data and research findings. This includes both collaborative research efforts and commercial transactions, allowing other entities to utilize the data and results for their work.	This objective will significantly enhance CIC bioGUNE's ability to leverage a broader range of health data sources, improving the quality and depth of research conducted by the Precision Medicine and Metabolism laboratory. By enabling seamless sharing of data and results, CIC bioGUNE can foster stronger collaborations and open up new revenue streams through commercial data transactions. This will ultimately increase the research output, innovation capacity, and financial stability of the center.
Ensure privacy, security, and traceability of	Implement robust measures to protect the confidentiality and integrity of health research data	Ensuring robust privacy, security, and traceability measures will build trust with data providers and research partners,

health research data and results	and results, ensuring compliance with regulations and maintaining detailed records of data provenance and usage.	reduce the risk of data breaches, and ensure compliance with regulatory requirements. This will protect CIC bioGUNE's reputation, avoid potential legal liabilities, and enhance the overall integrity and reliability of the research data handled by the Precision Medicine and Metabolism laboratory. Maintaining high standards of data governance will also facilitate smoother audits and reinforce the center's commitment to ethical research practices.

3.4.5 Actors involved

Table 11 UC3 actors involved

Actor	Business area	Type of impact	Description of the impact
Data integration and analysis specialists	IT	Direct	Data integration and analysis specialists will be responsible for integrating, analysing, and preparing new data sources. They will benefit from improved data quality, accessibility, and more efficient research workflows, enabling them to conduct high-quality and accurate research.
IT Security Specialists	IT	Direct	IT security specialists will be responsible for implementing and maintaining enhanced security and privacy measures. This will include managing encryption, access controls, and monitoring systems to ensure data integrity and compliance with regulations.
Research project managers	Management	Direct	Research project managers will oversee the integration of new data sources, the implementation of data sharing protocols, and coordination between departments and external partners. They ensure smooth data flows, adherence to timelines, and alignment with research objectives.
External data collaborators	Others (External Partners)	Indirect	External data collaborators, including data providers and research partners, will benefit from secure data sharing agreements and enhanced data privacy measures. They will have improved access to high-quality data and research results, fostering collaboration and innovation.
Compliance officers	Legal	Direct	Compliance officers will ensure that all data handling and sharing practices comply with relevant regulations and standards. They will be involved in auditing data processes and maintaining compliance

			documentation, and ensuring adherence to privacy laws.
Financial administrators	Accounting	Indirect	Financial administrators will manage the financial aspects of the data integration and sharing initiatives. They will handle budgeting, funding allocation, and financial reporting related to these projects, ensuring proper financial management and accountability.

3.4.6 User & system requirements

Table 12 UC3 requirements

Business objective	Requirement
Facilitate integration and sharing of health research data and results	Develop a centralized data repository for storing integrated health research data and implement data integration workflows to automate data ingestion from various sources
	Implement APIs for seamless data sharing with external partners
	Establish data governance policies to ensure data quality and consistency
	Develop user-friendly data visualization tools for researchers
	Provide training programs for researchers on new data sharing protocols and tools
Ensure privacy, security, and traceability of health research data and results	Implement advanced encryption protocols for all data
	Develop a comprehensive access control system to manage user permissions
	Develop training programs on data privacy and security protocols
	Implement data anonymization techniques to protect sensitive information
	Set up a system for tracking data usage, sharing metrics, and maintaining version histories to ensure data traceability.

3.4.7 Expected benefits and KPIs

Use case 3 expects the following benefits:

- The process of sharing health research data and results will ease collaboration and the access of information and knowledge between researchers.
- Enhanced privacy, security, and traceability will ensure that health research data is protected and secure, while also maintaining traceability to ensure accountability and transparency.

- Advanced data practices: PLIADES promotes sustainable data practices, which can help to improve the quality and reliability of health research AI models.
- Accelerated innovation: By facilitating data sharing and collaboration, PLIADES can help to accelerate the pace of innovation in health research.

The following KPIs will be measured to evaluate the implementation of the objectives and consequently the level of achievement of the expected benefits.

Table 13 UC3 benefits and KPIs

Indicator	Description	Current value	Future expected value	Expected date of achievement
Data integration efficiency	Measures the time required to integrate new health data sources into the centralized repository	24 h per data source	3 h per data source	During the implementation
Data sharing frequency	Tracks the frequency at which data and research results are shared with external entities	2 datasets/month	10 datasets / month	Before 6 months after implementation
Automated data quality checks	Measures the number of automated data quality checks performed per data source to ensure adherence to governance policies	-	10	Before 12 months after implementation
Researcher satisfaction with data visualization tools	Measures the satisfaction of researchers using data visualization tools through surveys	-	90%	Before 12 months after implementation
Training program completion rate	Measures the percentage of researchers who have completed training on new data sharing protocols and tools	0%	90%	Before 6 months after implementation
Encryption protocol implementation	Measures whether advanced encryption protocols have been implemented for all data.	Partial	Full	End of the implementation

Access control implementation	Measures the extent to which a comprehensive access control system has been implemented to manage user permissions	Partial	Full	Before 24 months after implementation
Privacy and security training completion rate	Measures the percentage of staff who have completed training on data privacy and security protocols	0%	90%	Before 6 months after implementation
Anonymization compliance checks	Measures the number of compliance checks performed per data source to identify data that may compromise anonymity	0	5	End of the implementation
Data traceability implementation	Measures the extent to which the system for tracking data usage, sharing metrics, and maintaining version histories has been implemented	None	Full	End of the implementation

3.5 Use Case 4: Integrating data life cycles of smart vehicles for CCAM operations and ADAS/AD functions

3.5.1 Use case overview

Use case 4 operates within the domain of Connected, Cooperative and Automated Mobility (CCAM) and Advanced Driver-Assistance Systems (ADAS)/Autonomous Driving (AD) functions. It emphasizes integrating data life cycles of smart vehicles to support the development and optimization of AI-based models for autonomous driving. The context spans industrial and mobility data spaces, involving edge-cloud pipelines and interoperability frameworks among them. It relies on data that originate from onboard vehicle sensors, fixed roadside infrastructure, alongside publicly available data. The main focus is to harmonize diverse data sources, enabling seamless data sharing, and maintaining data sovereignty while supporting innovation in ADAS/AD systems. This use case will bridge the gap between the mobility and the industrial domains, providing insightful data between organizations that perform R&D for ADAS functionalities and industrial automotive developers, who develop industrial ADAS products to be integrated with automotive end-user companies.

The use case will occur among various European countries, where various stakeholders are located and are interested in the adoption of Data Space infrastructure into their activities. Their coordination and cooperation in the development of such CCAM solutions can greatly assist the improvement of large-scale mobility and industrial applications. The utilization of data from diverse and completely heterogeneous conditions, and their efficient sharing, will highly improve the potential for greater engagement and collaboration between automotive providers, ADAS providers and research organizations across Europe, fostering the advancements of mobility and automotive industry.

The overarching goal of use case 4 is to create an advanced data integration and sharing framework, which will allow the overall enhancement of ML/DL methods for ADAS functions, while also providing access to larger amount of data for improving autonomy, traffic management and energy efficiency. Main aspect of use case 4 is also the assurance that data sharing is performed following all the regulations and privacy guidelines, leading to trustworthy data exchange between different types of stakeholders and providing benefits to everyone involved.

3.5.2 Background

3.5.2.1 Participants

The partners involved in this use case include research organizations, universities and companies, who are directly related to the domain of mobility, develop ADAS/AD applications, while also their applications can be useful for the automotive industry.

The Centre for Research and Technology Hellas (**CERTH**) is a research organization working on various topics related to mobility, including the development of perception systems for autonomous driving, navigation and localization systems for outdoor systems. CERTH with their own self-driving vehicles will provide valuable data sources to PLIADES, while also aims to improve the overall efficiency of the developed AI methods, through the utilization of more available data.

Taltech and the autonomous vehicles research group is working on research, development, and implementation of autonomous vehicles, industrial mobile robotics, and smart city testbed. In particular, they work on autonomous driving technologies for self-driving vehicles, simulations and validations in terms of safety, usability and edge case detection, mapping and virtual environment creation based on LiDAR and camera data from UAVs, smart city testbed development and user experiments in the future urban environment.

UC3M is an internationally recognized academic institution, noted for its commitment to research in multiple areas, such as robotics, mobility and electronics, among others. Within the framework of this project, UC3M's participation focuses specifically on the field of mobility, with a particular focus on advanced driver assistance systems (ADAS), autonomous vehicles and associated vision and sensing technologies. The university will contribute its expertise in this field, as well as the necessary resources and systems to ensure an efficient and rigorous development of the planned activities.

All three organizations perform extensive research on applications of autonomous vehicles, aiming to develop AI algorithms that can perform efficiently in various weather and lightning conditions for tasks, such as:

- Road segmentation and road line detection
- Traffic sign understanding
- Pedestrian and vehicle detection and tracking
- Pedestrian behaviour prediction

The Fundación Centro de Tecnologías de Interacción Visual y Comunicaciones (**VICOMTECH**) is a non-profit applied research centre, based in San Sebastian (Spain), established in 2001, and specialized in interactive computer graphics, multimedia, artificial intelligence and data processing. VICOMTECH cooperates with various local, national and European organizations to advance technology in multiple fields, such as Connected, Cooperative and Automated Mobility (CCAM). VICOMTECH is actively working and coordinating in numerous CCAM projects, in the context of EU-funded frameworks like Horizon Europe, where VICOMTECH researches on vehicle connectivity, perception, and automation, contributing to safer, more efficient, and sustainable transportation systems across Europe.

BasqueCCAM is an association that aims to position the Basque Country in the field of CCAM by leading and structuring the knowledge and capabilities of local public and private organizations. The organization is currently spearheading the implementation of procedures and methodologies to validate vehicular communications across diverse scenarios at their Living Lab. BasqueCCAM intends to provide the PLIADES Use Case with data sources derived from the ongoing testing at their Living Lab.

Both the research organization of Vicomtech and the association of BasqueCCAM carry out applied R&D in the domain of automated vehicles and aim to develop high-end AI-based methods that can be potentially integrated with commercial vehicles. Indicative scenarios and applications that are related to the aforementioned partners are:

- Automated parking
- Traffic sign recognition
- Situation awareness
- Intelligent perception

ZERO GmbH develops a holistic and innovative "Connectivity & Visualization Module“, called „CoViMo", which can also be used and expanded across industries due to its modularity. This module covers functions such as over-the-air updates, online fleet management, remote diagnosis, maintenance and control of vehicles and machines. ZERO plans to develop a semi-autonomous vehicle for logistics purposes. To this end, a car-following model has been developed to specifically address this purpose, but requires further validation with real data and the relevant variables and influencing factors need to be determined.

AVL List GmbH is a global leader in the development, simulation, and testing of powertrain systems, including advanced technologies for electric, hybrid, and internal combustion engines. With decades of expertise in innovation, we are now collaborating with focus into the automotive data space. Drawing on the deep knowledge and collaboration within the automotive cluster, AVL leverages industry insights and partnerships to address real-world challenges and drive innovative solutions. AVL develops ADAS functions in production for their clients who belong in the automotive industry.

CVUT is an internationally recognized university of which the participating Faculty of Transportation Sciences works on the short-term prediction of traffic using traffic-flow related data, providing information on the current state of the urban traffic network and the current state of the public transport in the areas of interest. Specifically, for selected parts of the urban traffic network, CVUT has access to data originating from traffic monitoring equipment with vehicle counters, video observations, occupancy detectors and speed-meters. Public transportation information that is used is related to the GPS position, delays and scheduled time of transportation means

CEIT cooperates with Gipuzkoa Living Lab (GLL) as a technical partner for traffic management. CEIT has deployed a range of detection and communication devices in real-world conditions, which are able to provide real-time data of V2X communications, RSUs, LiDAR’s and cameras, leading to the research and development of state-of-art methods for traffic coordination.

The research centre **PATRIC** provides works in the interaction of automated minibuses with other vehicles using enhanced floating car datasets (location, speed, wheel slip, , weather, pedestrian / cyclist detection, risk situation detection) and on-board analysis of the environment, which include geo-located road quality and traffic signs. These data can be provided to traffic management systems, owned by regional road maintenance organizations and regions, support the vehicle routing processes.

3.5.2.2Data Spaces Readiness

VICOMTECH has expertise in the creation and utilization of Data Space technologies, particularly specialized in Connected, Cooperative, and Automated Mobility (CCAM) contexts. Currently, VICOMTECH is leading the development of the FAME project's Test Data Space (TDS), which is expected to be delivered by mid-2025. This initiative aims to provide a robust and secure data infrastructure to support data sharing for CCAM projects and services.

BasqueCCAM has recently joined a national project to further develop its expertise in diverse mobility data space use cases. The organization will concentrate on the governance and scalability of a use case that establishes a global, interoperable data space. This use case will prioritize data governance and the interaction between various stakeholders, promoting collaboration among local authorities, multimodal transport operators, and other public and private entities within the ecosystem.

AVL List GmbH is making good progress in its readiness in the automotive data space, leveraging practical insights from multiple proof-of-concept (POC) projects with different customer use cases. These projects allow AVL to build deep expertise, refine our approach, and address real-world challenges in data-driven collaboration. By continuously learning and adapting, AVL aims to enable support, secure, and efficient data exchanges within the automotive ecosystem, creating value and driving innovation for everyone involved.

PATRIC has started its basic experience related to Data Space topics in 2023 together with CVUT by testing new concept of storing and manipulating road traffic services data. This concept was further elaborated by PATRIC and currently is under development of future product.

The rest of UC4 involved partners have not any prior experience with data spaces. However, there is extensive experience in the development of mobility technologies, along with AI for ADAS functions.

3.5.2.3Current Business Scenario

Two business scenarios have been identified within Use Case 4 – "CCAM operations and ADAS/AD functions in smart vehicles". The first one refers to the development of AI-based technologies that are developed, either from industrial companies or research organizations, in order to advance the existing state-of-art on vehicle perception and allow automated and autonomous vehicles to operate in various weather and lightning conditions. The second business scenario is related to the traffic management and coordination of connected and automated vehicles, which is performed through AI-based methods, relying on great amounts of data from diverse traffic environments.

On the one hand, these scenarios include partners who belong to the mobility domain and perform in-house R&D for ADAS systems and traffic management. On the other hand, these scenarios also include partners who provide ADAS functions to production and their clients, as a result of the performed R&D processes.

3.5.2.3.1 AI-based ADAS development

Development of AI methods for ADAS applications is performed by various stakeholders around Europe. In order to perform R&D on these topics, each partner uses their own self-generated data from their own autonomous vehicles, or rely on publicly available datasets that are directly related to the topic of interest. However, currently in order to utilize multiple datasets from different locations or from vehicles with different sensors, there is a significant number of pipelines that are needed to post-process the data and make them suitable for research and development purposes.

From the perspective of industrial domain, automotive ADAS providers require a mass amount of data, from simulated data up to real ones, in order to perform the requires validation cycles that lead to the final integration into the outcome products.

The scenarios of interest are based on multiple and different types of sensors (cameras, LiDAR, RTK-GNSS, etc.). However, each partner relies their own AI-development efforts on their self-collected data, using their own autonomous vehicles, as well as publicly available data that are available for this kind of applications. These datasets usually have major differences in the data format, the data collection conditions and the environmental features, creating a major impediment in the rapid utilization of multiple datasets for a common task. In addition, it becomes hard to generalize AI under diverse weather and lightning conditions (snowy, rainy, foggy, cloudy) and different road network settings (urban, rural, structured and unstructured).

Overall, the scenario of AI-based ADAS development involves plenty of data sources, that currently they might be represented with a different format by the related partners. These data are also connected with the corresponding label types that will be used for the development of AI-based methods.

3.5.2.3.2 AI-based traffic management

Development of AI-based methods for CCAM applications and traffic coordination usually relies solely on data that originate from specific regions, using the self-owned infrastructure of each partner. Specifically, among the partners of PLIADES, various of them have access to valuable data sources of traffic infrastructure from different European regions, however this knowledge is not shared between them. The scenario of AI-based traffic management involved various data sources that currently might be represented with a different format by the related partners. These data are also related to specific labels that will be used to enhance the AI-based traffic management methods.

3.5.2.3.3 Current Data Processes

The data processes that are common across the involved partners of this scenario are the following:

- Data collection
- Data filtering
- Data synchronization
- Data annotation
- Data sharing

Data collection refers to the production and storage of raw multimodal sensor streams from smart vehicles and static infrastructure. Currently, each partner collects data from their own infrastructure and vehicles, leading to data sets that consist of data originating from various sensor setups and hardware configurations, as have been previously described in the business scenarios as data sources.

This process starts from the streaming of sensor measurements and completes with the storage of these data into a server. This currently leads to mass amount of data, making it challenging to select the most useful parts of the collected data with the optimal quality that will assist the AI development.

During the data creation step, the format of the generated data is not the same among the involved partners. For example, ROS-based data are different from JSON-based datasets, but their information represents the same type of information. The format of data depends on the hardware manufacturer and the overall infrastructure deployed. This leads to inconsistencies among the generated datasets.

These data are saved in local servers of the data provider without considering any sustainable practices, leading to massive incoming data and potential burden when shared.

Data filtering refers to the execution of processes that improve the data quality by removing duplicates, excluding low-quality or noisy frames (e.g., blurred images), removing corrupted frames and generally eliminating irrelevant information that could lead to potential confusion in the system. Currently, data filtering is performed offline as part of the data pre-processing steps. This process involves manual reviewing of the collected data by human experts, followed by manual data exclusion in cases that it is needed.

Data synchronization refers to the process of establishing consistency between the data originating either from different sensors on the same agents or from different agents. So, vehicles equipped with multiple sensors are heavily reliant on the accurate synchronization of the sensor streams, so that all of them can be used for processing with accurate measurements. Also, especially in the case of AI-based traffic management, the synchronization of sensor streams from multiple vehicles and static infrastructure is necessary. The synchronization is based on both timestamps and geo-referenced data that are incorporated as metadata during the data collection. Currently, there is not a foundation to follow for data synchronization, while is also directly connected to the data format and the outcomes of the Data Collection process.

Existing tools, such as message filters provided for ROS-based infrastructure, or V2X communication protocols provide such functionalities. However, data synchronization is not always performed and many systems rely on a single data source and there are limitations that arise from supporting standardized data formats and data sources with differences in the operating frequency.

Data annotation refers to the process of generating labels for the collected data. These labels might refer to various AI-based tasks, such as object detection and semantic segmentation, or might be the recognition of events. The format of the labels can be either visual (image masks), text description or even audio labels. Also, the annotation can be related to different types of data and can have different formats for each one, e.g., the labels for point cloud data can be different that the label for image data.

Currently, the data annotation process is performed differently among the different partners of PLIADES. On the one hand, there are manual annotation procedures using some publicly available tools, either online or offline to extract the corresponding labels. On the other hand, some partners such as VICOM and BCCAM already use some semi-automatic AI-based labelling pipelines that are executed in the cloud; however, they always need human verification.

Another aspect of the data annotation process is the format of the labels. Currently, there is not standard followed by all partners and there are several available formats that are used by AI-based methods, such as YOLO format, COCO format or ASAM OpenLabel[9].

Data sharing refers to the data interchange among various partners who share common needs and would benefit from having access to data from different conditions, sensors or infrastructure.

Currently, data sharing is not an automated process and also requires great effort to adapt the available datasets, so that they will be usable. From the perspective of data providers, existing data sharing practice is to upload the generated datasets on a publicly available server, so that everyone has access to the data or to private servers, where access is selectively provided. In both approaches, this process requires direct contact between the two parties (provider and consumer).

On the side of data consumers, the data are downloaded locally to a computing unit, their integrity cannot be guaranteed, while also improvement by the data provider cannot be performed online, but require complete transfer of the dataset. Also, the data consumer needs to manually search for the available datasets, through online catalogs and select the optimal dataset, based on samples provided and their description.

3.5.2.4Pain Points

Both business scenarios contain some common challenges related to the performed data operations that will be described below. First of all, a great challenge is related to the accuracy and

[9] https://www.asam.net/standards/detail/openlabel/

representativeness of the edge-level computed data. This includes the processes of data filtering and the ground truth validation. Data quality also affects these processes, which can be hindered by: i) inconsistencies in data format; ii) completeness of data due to human error; and iii) communication overhead caused by the pre-processing steps prior to integration. Another critical challenge related to the data interoperability that occurs from the complex processing pipelines and the multiple data formats that are used. This could be resolved through the utilization of open formats, globally accepted from multiple stakeholders, such as the OpenLabel. The computation scalability arises more challenges, since various tasks, such as the privacy masking, require online computing and therefore high-performance hardware to perform such operations, on the edge or on the cloud. Human interaction with the AI application also creates important challenges, through the designed web applications for labeling inspection and streaming of data among multiple locations.

3.5.2.5Barriers

3.5.2.5.1 Legal Framework

The mobility and the industrial sectors are subject to several regulations that ensure safety, data protection and environmental responsibility, especially for the integration of AI technologies and automation. These regulations, along with the ones presented in Section 3.1.1, guide the adoption of advanced technologies for operational and compliance decisions:

- Regulation (EU) 2022/1426: lays down rules for the application of Regulation (EU) 2019/2144 of the European Parliament and of the Council as regards uniform procedures and technical specifications for the type-approval of the automated driving system (ADS) of fully automated vehicles. It provides the guidelines for type-approving three use-cases-oriented systems:
    - Fully automated vehicles for the transport of passengers/goods within a predefined area (e.g., robotaxis).
    - “Hub to Hub”: Fully automated vehicles for the transport of passengers/goods in a predefined route.
    - “Valet parking”: Fully automated mode for parking in predefined installations.

The adoption of AI technologies in the mobility and industrial domains requires the adherence to specific standards and specifications, in order to ensure safety, interoperability and performance reliability.

- ISO/SAE 21434, Road Vehicles – Cybersecurity Engineering: This standard addresses cybersecurity in road vehicles, ensuring that connected and automated systems are designed, developed, and maintained with robust security measures. It covers threat and risk assessment, cybersecurity design, and vulnerability management across the vehicle’s lifecycle. Compliance with ISO/SAE 21434 is crucial for manufacturers and developers aiming to implement secure AI systems in vehicles.

3.5.2.5.2 Social Framework

The social framework for use case 4 in PLIADES identifies the key societal elements and challenges influencing the integration of AI-based technologies in CCAM operations and ADAS/AD functions. These factors include social relations, equality and diversity, privacy, security, health, sustainability, globalization, and cultural considerations. The social frameworks are described in Section 3.1.2.

3.5.2.5.3 Economic Framework

The economic framework for use case 4 in the PLIADES project focuses on the financial aspects that shape and are influenced by the integration of AI-based technologies for CCAM operations and

ADAS/AD functions. This framework highlights how these advancements contribute to cost efficiency, resource optimization, market opportunities, scalability, and long-term economic impact.

AI-driven solutions in smart vehicles and traffic management offer significant cost-saving opportunities. Automation of data filtering, synchronization, and annotation processes reduces reliance on manual labour, lowering operational costs. Additionally, sustainable practices such as data reduction and optimized edge computing minimize energy expenses. Predictive analytics also play a role by enabling proactive system maintenance, reducing downtime and the associated financial burden. The integration of standardized data protocols and AI-based methods enhances resource allocation. By reducing redundant data processing and storage efforts, stakeholders can optimize their use of computational and physical resources. Efficient data-sharing mechanisms further reduce duplicated effort across collaborators, while seamless workflows improve the utilization of high-performance computing infrastructure, supporting scalability.

The adoption of AI technologies creates new market opportunities in mobility and industrial domains. Improved ADAS functionalities and enhanced traffic management systems drive innovation and attract investment, generating new revenue streams. Technology providers supplying data platforms, annotation tools, and AI models benefit from increased demand, while automotive manufacturers leverage these advancements to deliver high-value autonomous vehicle products.

Economic benefits are further amplified through scalable and interoperable systems. Efficient data-sharing mechanisms enable stakeholders to maximize the utility of shared datasets while reducing acquisition costs. Solutions developed under the PLIADES framework are designed to be adaptable for cross-industry applications expanding their market reach.

While regulatory compliance, including adherence to GDPR and the EU AI Act, incurs initial costs, these measures mitigate financial risks such as penalties and reputational damage. By ensuring ethical and lawful data usage, the PLIADES framework builds trust among stakeholders and strengthens market confidence.

Long-term economic impacts include fostering innovation and job creation in technical fields such as AI development and data engineering. Enhanced public infrastructure efficiency, enabled by advanced traffic management and autonomous systems, reduces economic losses from congestion and accidents. Additionally, these developments position European industries as global leaders in AI-driven mobility technologies, enhancing their competitiveness on the international stage.

Finally, the economic viability of the PLIADES framework is reflected in its return on investment. Faster development cycles, reduced redundancy in data collection, and more accurate AI models contribute to higher adoption rates and improved market penetration. By aligning technological advancements with financial sustainability, the PLIADES framework ensures economic growth and long-term value creation for all stakeholders involved.

3.5.3 Future Scenario

All involved partners will continue to collect data and develop AI algorithms using their own infrastructure; however, they will also be able to utilize data from other data providers in a direct and efficient way, without adding any additional overhead to the development process. In addition, they will be able to share their data with the option of gaining financial benefits, without affecting the data ownership and the usage rights. The creation of a common protocol for collecting data will create a guideline regarding the data format and the data sets' content, enabling the direct utilization of various datasets. Then, the utilization of automated and semi-automated annotation pipelines with human knowledge injection mechanisms will speed up the process of data preparation, prior to AI training. Finally, the usage of data will be based on privacy and legislation-aware processes, allowing the direct usage of data products by the data consumers, while simultaneously adhering to all privacy and ownership constraints. Below a complete description of each separate business scenario will be provided.

3.5.3.1 AI-based ADAS development

Research organization, who tend to produce public datasets as an outcome of their research efforts, will follow pre-defined guidelines during the data collection process. Then, they will be able to utilize automated procedures to filter the data and potentially add annotations related to the topics of interest. These data will be available through AI-based data brokering schemes, allowing the direct access to data by every potential stakeholder.

Industrial companies will be able to use data originating either from research organizations or other industries in order to improve the development of ADAS products. Significant aspect for industries remains that of data sharing, due to privacy and policy constrains of the clients. Through PLIADES, the future scenario of ADAS development will allow each company to control the access rights and to whom the data are available, without any potential danger of privacy compromise.

3.5.3.2 AI-based traffic management

Traffic coordination and management will highly benefit from the introduction of PLIADES technologies into the overall pipeline of technology providers. All involved partners, irrespectively of the hardware used for the RSUs and road sensors, will follow a common protocol to collect and process their data. Also, they will all have access to shared data from similar infrastructure, providing them the ability to cover more cases and scenario that happen in real-world cases. Privacy concerns will be automatically resolved by the automated pipelines, while data access will be always controlled by the data owners, limiting the unwanted access of data. The outcome of this future scenario will be the improved and faster development of traffic management methods, covering more edge-cases and real-world scenarios for diverse weather and environmental conditions.

3.5.3.3 Future scenario data processes

The processes that will be affected by the PLIADES framework are relevant to both business scenarios that belong to the use case 4. To this end, the following processes will refer both to the “AI-based ADAS development” and the “AI-based traffic management”. The business processes are directly related to the data operations, which were mentioned in the previous section, and that are performed either by data providers or data consumers or both.

Data collection’s evolution is related to the standardization of data formats that are used by the involved partners, as well as to the introduction of sustainable practices, leading to the green approaches for data collection.

The format of the collected data will be based on standardized formats that are widely used in mobility and industrial applications. Such formats are common file formats that are widely used for images, sound, videos, etc. such as .JPG, .MP4, .PNG, etc. Also, data collection will use a common format based on XML structures to describe the relative placement among different sensors and also any other necessary metadata that can assist the time and location synchronization of different data source.

Sustainable practices will be established in order to reduce the amount of data collected, but without compromising the information available. These AI-based methods will be executed either real-time, during the streaming of sensor data to a computing unit, either offline, as a post-processing step to the collected data.

Through the PLIADES framework, **data filtering** will undergo important improvements that will lead in improved performance of these processes, while also reducing the required time. Specifically, data filtering will be performed online, on the edge, by performing a sequence of filtering pipelines in a specified order. The outcome generated will be an updated version of the data with increased quality, along with quality metrics for various types of data.

These pipelines will evaluate the data quality and will automatically discard data that do not contribute in the collected data sets. These processes will be common for any time of data, either originating from floating vehicles' sensors, static infrastructure, or onboard analytics from vehicles. AI-based approaches will be developed as part of the overall data processing pipeline and will take into account data quality aspects to perform automated data filtering. In addition, specific modules for privacy-related filtering will be developed, in order to either discard frames or edit them accordingly, in order to follow the privacy guidelines. Such methods include the blurring of faces in image data and other potential needs that will be extracted from the related tasks.

Data synchronization will be evolved through the PLIADES framework in a way that will support also the rest of the developed architecture. Specifically, the developed modules will support the data formats that will be used by the data collection process and will automatically run synchronization pipelines, whenever is available. This will lead to more complete data sets, which will contain information from multiple agents and infrastructure, capable to be used for AI-based processing in later stages.

Also, multiple data streams from the same agent will be synchronized using automated techniques based on time protocols, while these methods will also cover the cases that there is a great different between the operating frequency of the streams. This business process will support all the available data types that have been previously described.

Regarding **data annotation**, the existing automated annotation methods will be further expanded with human knowledge injection methods.

Also, annotation will be based on standardized formats, such as the ASAM OpenLabel, which can cover the needs of both object-level and action-level annotations, while the labeled classes will be defined following the Automotive Global Ontology (AGO) [9]. Data annotation will also include the handling of metadata insertion, suitable for the improved discoverability of data. To this end, automated annotation tools will be developed, while the Digital Logbook by AVL will be expanded to cover more needs of the industrial and the mobility domains.

Data sharing will undergo significant improvements by the incorporation of data spaces and the PLIADES framework. The previous data processes will prepare the data in a way that will allow the seamless exchange among various partners, allowing them to directly use them.

From the perspective of data providers, the data sharing process uses as input the data that have been generated by the previous processes of Collection, Filtering, Synchronization and Annotation. These data will be extended with semantics metadata that will contain all the essential information regarding the ownership of data, the access rights and the potential regulations of usage. These data with their metadata will be available to the AI-based brokering schemes and provided through the developed AI-based connectors. Data providers will also be able to actively maintain and update their provided datasets.

From the perspective of data consumers, they will first of all make use of the AI-based querying system and the AI-based brokers in order to identify all the available datasets within multiple dataspaces and receive an accurate suggestion on a dataset, according to their needs. Data consumers will also agree on the pre-defined policies for the data usage, as part of their connection on the dataspace, ensuring the proper use of data

3.5.3.4Cross Data Space connections in the Use Case

Based on the above-mentioned data types, the contribution of each partner in the scenario of AI-based ADAS development is described below, relating both to the mobility and industrial data spaces.

CERTH will act as a **data provider** and **data consumer** to the **mobility** data space for data related to the perception of autonomous vehicles. Specifically, will provide data originating from their autonomous vehicles (LiDAR, RGB images, location information, vehicle velocity, acceleration). For these data, CERTH will also provide annotation labels for the tasks of 2D and 3D object tracking, segmentation masks, as well as behaviour labels regarding the intention of drivers. CERTH will also consume the same types of data that might originate from other data providers, in order to develop AI algorithms for AI-based vehicle perception.

Taltech will act as a **data consumer** by the **mobility** data space for developing and boosting the performance of Ai-based perception methods for understanding of road semantic element, such as traffic signs, road marking, lanes, pedestrians, vehicles and their behavior. Also, Taltech will work on the generalization of AI under diverse weather and illumination conditions and under different environmental settings. These data include LiDAR measurements, RGB images, IMU and vehicle state measurements.

UC3M will act as a **data provider** to the **mobility** data space with data suitable for the development of AI-based perception system for understanding key road elements, such as traffic signs, vehicles, pedestrians, lanes and their state. These data will originate from their autonomous vehicle operating in the metropolitan area of Madrid and include LiDAR scans, RGB images, GNSS and IMU measurements. Also, will provide annotations for 2D and 3D object detections and tracking.

VICOM will act as a **data provider** to the **mobility** data space for raw data and vehicle state data originating from their autonomous vehicles. Specifically, these data include LiDAR measurements, RGB images from fisheye cameras, location information from GPS and CAN bus data. These data can be useful for the development and the enhancement of AI-based perception methods in autonomous driving. VICOM will also provide annotations regarding the 2D and 3D detections, road segmentation masks and events.

BCCAM will act as a **data consumer** from the **mobility** data space with data from sensor-equipped vehicles in order to aid the comprehension of road semantic elements and the overall improvement of the autonomous navigation functions that BCCAM develops. Specific scenarios, such as automated parking, traffic sign recognition and obstacle detection, will be enhanced, for diverse weather and lightning conditions, as well as different road network settings.

ZERO will act as a **data consumer** from the **mobility** data space in order to develop their semi-autonomous vehicle for logistics purposes. Specifically, ZERO needs to have access to plenty of data that will enable the development of an AI-based car following model and requires validation with real data, in order to replicate the human-like behavior in the vehicle.

AVL will act as a **data consumer** from the **mobility** data space, in order to develop functionalities that are interesting for the **industrial** data space. Specifically, AVL will use data coming from various vehicles, include sensor measurements, vehicle state and environmental conditions, in order to develop their own ADAS products for the automotive industry. In addition, AVL will act as a **data provider** to the **industrial** data space, with data related to the dynamics, the maintenance and the state of the vehicles.

The contribution of each partner in the scenario of AI-based traffic management is described below, showing their relation to the **mobility** data space.

CEIT will act as a **data provider** and **data consumer** to the **mobility** data space for traffic related data. Specifically, CEIT can provide sensor data and environmental data originating from fixed LiDARs and cameras located in specific road points, along with V2X communication information from ITS-5G modems, providing information for the state of traffic in specific locations and the environmental conditions. Also, CEIT will consume data from static sensors, intersection detectors and annotated data from fixed infrastructure to improve the understanding and identification of different traffic road elements in different conditions and regions.

CTU will act as a **data provider** of traffic-flow related data to the **mobility** data space. These data will provide information about the current state of the urban traffic network and the current state of the public transport in the area of interest. These data will originate from intersection detectors, optical cameras in observation points, inductive loops and speed measurements, along with the publicly available public transport information, capable to improve the short-term traffic predictions in traffic scenarios.

PATRIC will act as both, a **data provider** and a **data consumer** for the **mobility** data space in the area of automated minibuses routing and their interaction with other vehicles in terms of fulfilling required task and minimizing risks, providing information in traffic management systems. In addition, PATRIC acts as a **data consumer** of data emerging from vehicle sensors, data from infrastructure (e.g., localized weather, emissions, road surface, and traffic information). These data will be analyzed and the outcomes will be **provided** to regional road maintenance organizations.

3.5.4 Objectives

Table 14 UC 4 business objectives

Business objective	Description	Impact
Enhanced data standardization and interoperability	Implement standardized data formats and protocols across all partners within the PLIADES framework. This objective focuses on creating uniformity in data collection, formatting, and storage processes. The standardization will facilitate seamless integration, minimizing compatibility issues and enabling more effective data sharing and use across different platforms and applications. It also includes a standardized approach to data annotation across all partners to ensure consistent labelling for training AI models, based on common standardized annotation formats (e.g., ASAM OpenLABEL).	This objective is essential for reducing friction in data exchange, lowering the need for data format conversion, and enhancing data interoperability across multiple systems. The adoption of industry standards will allow faster integration of data from various sources, supporting more comprehensive and reliable data-driven insights.
Sustainable and cost-efficient data operations	By automating data filtering, synchronization, and processing workflows, this objective aims to reduce the time, labor, and resources currently required for data handling and preparation. Advanced data processing techniques, like edge computing and AI-driven filtering, will allow near real-time data handling with minimal manual intervention. Also, the automated filtering methods will reduce the amount of data collected, but without compromising the information available.	This objective will lead to greener approaches in the data storage, reducing the needs for large amounts of data, by the elimination of data that do not provide useful information or compressed data that lead to less storage needs. Automation will lead to significant cost savings and reduce dependency on human resources for data management. Faster data processing also improves

		response times, enabling near real-time insights and supporting more dynamic applications, such as AI-based traffic management and ADAS systems.
Improved data quality and annotation consistency	Implement automated data quality estimation methods that will be executed during the data collection processes or further operations. These AI-based quality extraction methods will operate on various types of data, such as images and point clouds, leading to higher-quality datasets. Quality assessment also refers to the annotations of the datasets, through semi-automated annotation processes that allow enhanced human knowledge injection methods, increasing the quality of the available annotations. A standardized annotation process will provide unified labels, enhancing model accuracy and training efficiency, ensuring high-quality labels.	This objective will lead to higher quality data, along with consistent and accurate labels, critical for the effectiveness of AI models, especially in ADAS and traffic management where precision is essential. Reliable labelling ensures that models can learn effectively from training data, leading to improved detection, classification, and prediction capabilities.
Secure and scalable data sharing and discoverability	Create a secure and scalable data-sharing infrastructure through data spaces that supports controlled data exchange among partners. This objective emphasizes data integrity, accessibility, and privacy while facilitating data sharing in real-time. The infrastructure will be equipped with AI-based data brokering capabilities, optimizing data recommendations based on user needs and data relevance.	Enhanced data sharing ensures that partners can access high-quality, up-to-date data without unnecessary delays, which is essential for applications like real-time traffic management and AI model development. A secure infrastructure also ensures that proprietary and sensitive data are protected, fostering trust and collaboration among partners.

3.5.5 Actors involved

Table 15 UC4 actors involved

Actor	Business area	Type of impact	Description of the impact
AI developers	IT	Direct	Design and train AI models to enable ADAS features like object detection, behavioral prediction, and traffic flow analysis. Contribute to achieving interoperability and standardization across datasets, ensuring generalizability of

			models for diverse conditions (weather, lighting, etc.).
Data engineers	IT	Direct	Implement advanced data filtering, synchronization, and annotation pipelines. Optimize dataset quality to enhance AI training accuracy and reduce data redundancy, aligning with sustainability goals.
IT administrators	IT	Indirect	Deploy and maintain secure and scalable data-sharing platforms, ensuring compliance with privacy laws (e.g., GDPR). Facilitate real-time data exchange and collaboration between partners in mobility and industrial domains.
ADAS integrators	Manufacturing	Direct	Integrate AI-based insights and models into vehicle systems for advanced functionalities (e.g., automated parking, lane keeping). Ensure robustness and adaptability of ADAS in complex and dynamic environments.
Traffic coordinators	Management	Direct	Use real-time traffic management insights generated from AI-based systems to optimize routing, reduce congestion, and improve overall traffic flow efficiency. Enhance coordination between automated and connected vehicles.
Policy regulators	Legal	Direct	Evaluate and ensure the use of AI technologies complies with EU regulations, including GDPR and the EU AI Act. Provide oversight for ethical AI deployment and data governance.
Research project managers	Management	Indirect	Research project managers will oversee the integration of new data sources, the implementation of data sharing protocols, and coordination between departments and external partners. They ensure smooth data flows, adherence to timelines, and alignment with research objectives.

3.5.6 User & system requirements

The user and system requirements for the PLIADES platform are directly related to the data processes that will be affected and advanced through the project. Each requirement is connected with the objectives that have been set for the Use Case and are described below in detail:

Table 16 UC4 requirements

Business objective	Requirement
	Definition and adoption of standardized data formats

Enhanced data standardization and interoperability	Definition and adoption of standardized metadata formats
	Definition and adoption standardized formats for annotation labels
	Automated conversion of existing datasets format into standardized formats
	Seamless data exchange between systems and devices through easy-to-use APIs
	Validation of incoming data against predefined schemas and standards
	A metadata repository must store details of standardized formats and protocols
Sustainable and cost-efficient data operations	Support of compression techniques to optimize data storage without losing critical information
	Automated identification and discard of redundant data
	Automated identification and discard of noisy and low-quality data
	Optimization of edge computing for on-device data processing and filtering, eliminating the reliance on centralized systems
	Automated deletion of outdated and expired data
	Monitoring of resource usage to optimize data operations
	Definition of policies for sustainable data operations and archiving
	Allow real-time processing of multimodal data
	Effective and scalable data operations
	Cloud-based data operations for cost effectiveness
	Integration of AI-driven insights and analytics into resource planning
Improved data quality and annotation consistency	Real-time quality assessment of multimodal data
	Automated monitoring of data adherence to privacy and quality standards
	Intuitive interface for semi-automated annotation processes
	Inject human knowledge during the annotation processes
	Provide feedback to annotator for refining AI-generated annotations
	Develop tools for real-time visualization of annotations
	Train annotators on the use of the automated tools
	Generate reports on the annotation quality and consistency of data

	Support metadata tagging in annotations to include context-specific information
	Define and establish numerical metrics to evaluate annotation quality and consistency
Secure and scalable data sharing and discoverability	Secure data-sharing infrastructure with encryption mechanisms
	Enable discoverability of datasets through AI-based brokering mechanisms and recommendation engines
	Ensure compliance with privacy regulations in data-sharing activities
	Role-based access control for data access
	Define data access policies to control data sharing
	Management of data sharing agreements and permissions through user-friendly interfaces
	Real-time anonymization of sensitive and personal data
	Continuous monitoring and logging of data-sharing activity

3.5.7 Expected benefits and KPIs

PLIADES will provide various benefits to the involved partners, which are connected to multiple aspects of the mobility and the automotive industrial domains. First of all, it is expected to lead in improved operational efficiency, as a result of the faster and automated data processing workflows and the reduction of manual intervention. It will also reduce the data storage costs, through the optimized data storage techniques and the resulting reduced resource allocation. In addition, PLIADES will benefit the robustness of AI models, through the utilization of improved and larger training datasets and methodologies, assisting the prediction and the capabilities of the AI models. PLIADES will also have a high impact on the data interoperability through the adoption of unified standards and the integration of seamless data sharing pipelines. Through PLIADES, the involved stakeholders will manage to achieve compliance with the privacy regulations, such as GDPR and similar regulations, avoiding legal risks and penalties. Finally, PLIADES will benefit in the cross-border collaboration, through the utilization of shared data standards and secure data-exchange platforms, aligning with the global market trends and policy requirements.

The Key Performance Indicators that will support the aforementioned expected benefits, are described below:

Table 17 UC4 benefits and KPIs

Indicator	Description	Current value	Future expected value	Expected date of achievement
Reduction in data processing time	Time saved in filtering, synchronizing, and	Depends on data types	>50% reduction	6 months post-implementation

	annotating datasets through automation	and dataset size		
Increased AI model accuracy	Improved ADAS functionality (e.g., object detection accuracy in diverse weather conditions)	Depend on the task and the input type	>5% improvement	12 months post-implementation
Reduction in data storage	Decrease in storage requirements due to filtering methods, sustainable practices, and compression methods	Depends on the dataset size and type	>40% reduction	9 months post-implementation
Compliance with data privacy standards	Alignment with GDPR and other privacy regulations	Partial	Full compliance	6 months post-implementation
Time to develop AI-based products	Faster product development cycle through improved workflows and more available data	Depends on the data available and the task	>30% reduction	9 months post-implementation
Increased data annotations	Automated annotation methods with human knowledge	Depends on dataset type and size	>50% improvement	6 months post-implementation
Increase in reusable datasets	More available datasets that can be re-used for a same task/project	Depends on the task	>50% increase	12 months post-implementation
Enhancement of data quality	Reduction of data errors and inconsistencies	Depends on the type of data	>10% quality increase	9 months post-implementation

3.6 Use Case 5: Integrating data life cycles of WEEE/batteries management and car parts manufacturing operations

3.6.1 Use case overview

The UC5 is aiming to integrate data life cycles of Battery Energy Storage Systems (BESS) developed, installed and managed by LIBATTION. The BESS are installed in the EU region, mainly in Switzerland. Life cycle data collection of BESS is necessary for monitoring and maintaining of the systems.

Within the initial scope of the UC5, two main types of information have been identified to be of crucial importance for the integration of life cycles data of batteries. The two types of data flow will be managed independently on the dedicated data space. These data flows will be briefly described hereafter as well as the goal they will achieve.

- Data flow -1: Battery Operation

BESS require detailed data monitoring during their operation. To assure correct functionality, more accurate state estimations, predictive maintenance and proper display of system behaviour and performance, it is necessary to continuously log data from the Battery Management System (BMS). For this scope, battery data including data read from sensors (voltage, temperature, current, etc.) as well as status of many variables of the BMS, will be continuously logged and stored on the data space. The data space will be the virtual location to allow easy, secure and protected exchanged of data, as well as data analysts' tools (such as battery analytic) between data provider and data users.

- Data flow -2: Battery Digital Twin

This use case focuses on improving and automating maintenance tasks for BESS and electric vehicle (EV) batteries. LIBATTION and AVL will provide CAD models and product specifications in a dataspace. SIPBB will use this data to develop digital twins and generate synthetic datasets supporting computer vision applications. The data will be included in the dataspace. Moreover, SIPBB will train and share the results of the AI models for computer vision applications.

3.6.2 Background

3.6.2.1Participants

Libattion AG is a Zurich-based company specializing in sustainable energy storage solutions. Founded in 2018, the company focuses on upcycling used electric vehicle (EV) batteries from the European automotive sector to create high-quality energy storage and management systems. This approach extends battery lifecycles, reduces resource consumption, and supports industry decarbonization efforts. Libattion's product line, known as "e-Racks," offers modular and versatile energy storage systems ranging from 97 kWh to 60 MWh. These systems provide services such as frequency control, peak demand reduction, and fast EV charging, making them suitable for supporting critical infrastructure and integrating renewable energy sources.

The Switzerland Innovation Park Biel/Bienne AG (**SIPBB**) is a private Swiss non-profit organisation that conducts and supports industry-related and primarily applied research and development. SIPBB's Swiss Battery Technology Center (SBTC) provides engineering services and focuses on research and development in battery and battery related technologies. Particularly in areas of reinforcement learning and robotics for automated battery de-manufacturing. Among others, the relevant aspects for use case 5 is, that this requires access to a vast amount of real-world data as well as the generation of synthetic data for model training and simulation.

AVL List GmbH is a global leader in the development, simulation, and testing of powertrain systems, including advanced technologies for electric, hybrid, and internal combustion engines. With decades of expertise in innovation, we are now collaborating with focus into the automotive data space. Drawing on the deep knowledge and collaboration within the automotive cluster, AVL leverages industry insights and partnerships to address real-world challenges and drive innovative solutions.

3.6.2.2 Data spaces readiness

UC5 involved partners' familiarity with data spaces is rather limited. SIPBBs expertise in the field of technologies that enable the implementation of data spaces is largely theoretical and very modest.

However, all are eager to contribute their expertise and integrate their processes into the data space framework.

3.6.2.3 Current business scenario

The current business scenario foresees limited data and knowledge sharing between LIBATTION and its partners. Moreover, when data are shared, no consistent methodology is used. This scenario does not promote cooperation with other businesses involved similar activities or additional potential partners.

From the SIPBB side, the automation of EV battery disassembly faces significant challenges due to the lack of data sharing from manufacturers, which hinders the development of effective automated solutions. As a result, disassembly processes are still performed manually, leading to safety risks for workers, inefficiencies, and time-consuming operations. Automation requires access to detailed data, and while sharing CAD models greatly facilitates the training of AI models, collecting such data is currently a labour-intensive process. Manual data collection, including reverse engineering of EV batteries, is not only time-consuming but also prone to inaccuracies and inconsistencies. If CAD is shared, high-fidelity digital twins of the batteries could be created, enabling the generation of synthetic datasets. These datasets could then be used to train AI models for object detection and pose estimation, ultimately accelerating the path toward fully automated and safer disassembly processes.

3.6.2.4 Pain Points

Currently the most significant issue is access to design files, CAD- files and general information on the batteries form the manufacturers, mostly due to the unwillingness of OEMs to share these files and information. Even to gain access to the relevant entities in the EV batteries fields is difficult. Moreover, regarding the data, difficulties can be observed in data labelling, incorrect interpretation of failures or events, insufficient assumptions and poor performance of the integrated system.

3.6.2.5 Barriers

3.6.2.5.1 Legal framework

The legal frameworks that affect use case 5, apart from the ones described in 3.1.1, are the following:

- <u>EU Battery Regulation:</u> Establishes the framework for the Battery Passport, mandating transparent sharing of battery lifecycle data (e.g., origin, critical material, performance, recycling).
- <u>Battery Passport Compliance:</u> Data systems must facilitate structured and secure integration of battery lifecycle data. As of now this is not a legal requirement, however, by 1st February 2027 i.e., M38, all EV, LMT and industrial batteries >2kWh on the EU market will require a battery passport.

3.6.2.5.2 Social framework

Sustainability is important in the battery sector due to its role in mitigating environmental impacts and supporting a circular economy. Batteries in applications like electric vehicles and energy storage, require materials such as lithium, cobalt, and nickel, whose extraction has significant ecological and social consequences. Sustainable practices, including efficient recycling and lifecycle management, reduce resource depletion, minimize waste, and lower carbon emissions. The EU Battery Regulation also mandates minimum content of recycled materials by 2030.

The above practices have a major social impact as they minimize hazards and are aligned with the Green Deal initiatives. Another aspect of circularity is the ability to use products and materials at their highest value. The goal is to enable processes like maintenance, remanufacturing, reuse, recycling or refurbishing that reduce energy consumption and production of scrap. In this particular use case, accurate estimations of the State of Health (SoH) for example, of the produced batteries could lead to the suggestion of the optimal post-use application for a specific battery module.

3.6.2.5.3 Economic framework

The economic framework is dictated for the non-profit research institutions. This limits traditional product-based revenue streams but enables access to public funding, grants, and collaborative projects.

3.6.3 Future scenario

The data space associated with use case 5 will provide a permanent and centralised platform for data sharing that will accompany each installed BESS throughout its life cycle.

Figure 10 shows a schematic flow of data, applications and access to the data related with battery operation in the related data space. The data from the BESS will be periodically logged onto the data space according to a unified dataset interface. Data will be available to be visualised and processed by whoever has got access. Third party entities could access the data space to offer their services for battery analytic and any other battery supplier could provide their data to benefit from those services.

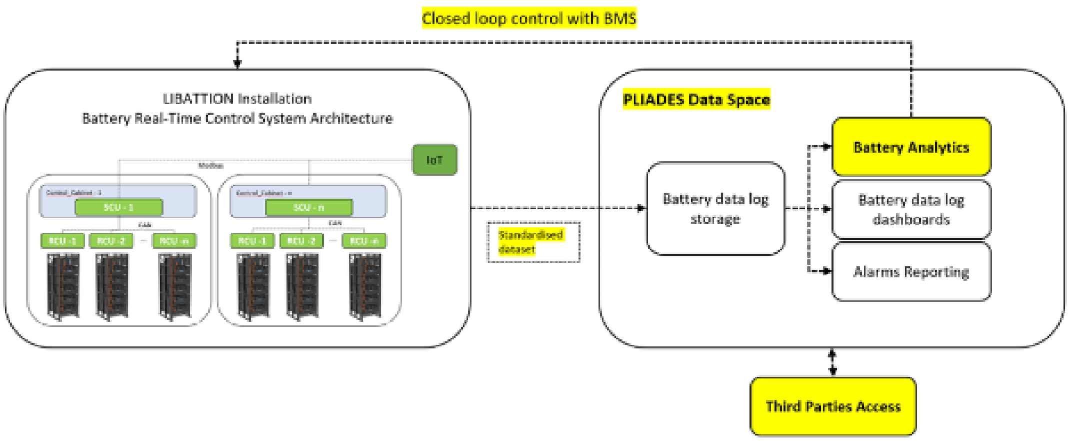

Figure 10 Future Business scenario for data related to Battery Operation

Figure 11 shows a schematic flow of data, applications and access to the data related with battery Digital Twin in the related data space. The battery digital twin developed by SIPBB is based on the required information shared by LIBATTION (such as battery CAD, data-sheet, bill of material, manuals, etc.). Once the digital twin of the battery is available on the data space, they will be available for supplier manufacturer and/or future entities dealing with the BESS (such as recyclers, second or third life applications, maintenance, etc.).

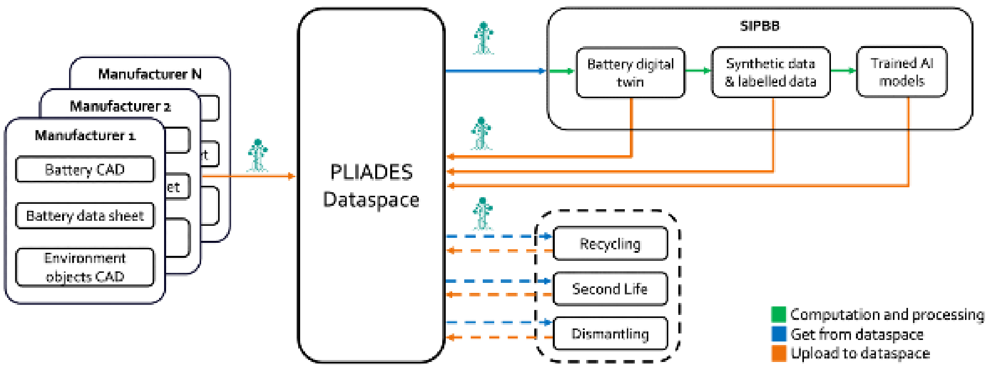

Figure 11 Future scenario for data related to Battery Digital Twin

In this way, the data space will act as a centralised platform where demand and offer for all topics related with BESS could meet.

For SIPBB, the implementation of the PLIADES platform, built on the concept of data spaces, offers a solution to the challenges currently faced in automating battery disassembly. By enabling manufacturers to securely share CAD models and associated data within a collaborative data space, the framework will streamline access to critical information needed for automation. This will eliminate the need for time-consuming manual data collection and reverse engineering, significantly reducing inefficiencies and inaccuracies.

With CAD data readily available in the dataspace, it will be possible to create accurate digital twins of batteries, facilitating the generation of high-quality synthetic datasets. These datasets can then be used to train AI models for object detection and 6D pose estimation, paving the way for automated and efficient battery disassembly. By leveraging the PLIADES ecosystem, stakeholders will benefit from a unified data ecosystem that accelerates innovation, reduces operational risks, and improves the overall efficiency of disassembly processes.

The application of federated cross-domain analytics for this use case is expected to focus on the optimization of the production process of Libattion, by incorporating advanced ML algorithms to accurately estimate the status of electronic batteries and their potential post-use strategy that heavily depends on that. The battery analytics that will be developed for that specific use case, will have cross-domain applicability and will be tested in diverse (but closely related) domains of the PLIADES data space. The Green Deal dataspace will be enriched with information (metadata) about available materials and post-use products, that through the use of PLIADES ecosystem will enhance circularity options.

3.6.3.1 Future scenario data processes

The future scenario will deploy improvements in many of the aspects of the existing data processes.

Data collection's evolution is related to the standardization of data formats that are used by the involved partners, as well as to the introduction of sustainable practices, leading to the green approaches for data collection. The format and the content of the data collected will be the minimal required to provide sufficient understanding of the status of the monitored systems.

Data filtering will provide improvements to the entire data flow process. Data filtering might be implemented at the data collection step, and at cloud level. The outcome generated will be an updated version of the data with increased quality and reduced storage size.

Data sharing will undergo significant improvements by the incorporation of data spaces and the PLIADES framework. The previous data processes will prepare the data in a way that will allow seamless exchange among partners, allowing them to directly use them.

From the perspective of data providers, the data sharing process uses as input the data that have been generated by the previous processes of Collection, Filtering, Synchronization and Annotation. These data will be extended with semantics metadata, that will contain all the essential information regarding the ownership of data, the access rights and the potential regulations of usage. These data with their metadata will be available to the AI-based brokering schemes and provided through the developed AI-based connectors. Data providers will also be able to actively maintain and update their provided datasets.

From the perspective of data consumers, they will first of all make use of the AI-based querying system and the AI-based brokers in order to identify all the available datasets within multiple dataspaces and receive an accurate suggestion on a dataset, according to their needs. Data consumers will also agree on the pre-defined policies for the data usage, as part of their connection on the dataspace, ensuring the proper use of data.

3.6.3.2Cross data space connections in the use case

In this use case, the dataspace serves as a central hub for data interactions between stakeholders. Manufacturers will upload CAD models of their batteries to the dataspace, providing the foundational data required for automation. SBTC will retrieve these CAD models and specifications to create high-fidelity digital twins, which will also be uploaded back to the dataspace. Using the digital twins, SBTC will generate synthetic datasets tailored for computer vision tasks, such as object detection and 6D pose estimation, and these datasets will likewise be contributed to the dataspace. Additionally, SBTC will train AI models using synthetic data and upload the resulting trained models, enabling other stakeholders to utilize them for automation and further research.

Data sharing is expected to take place in the PLIADES platform. Each data space participant can enter one or multiple dataspaces (e.g., Energy and Green Deal) possibly by installing an IDS-compatible connector software. Data and potential trained model exchange will be achieved through these connectors. At a second level, interconnection between multiple dataspaces will be achieved. An AI-broker will be in the position to perform a search to all of the PLIADES data spaces and suggest suitable AI implementations and datasets based on the keywords requested. The outcomes of the federated cross-domain analytics for use case 5 will ultimately be enhanced with training data and metadata from the whole PLIADES ecosystem.

3.6.4 Objectives

Table 18 UC 5 business objectives

Business objective	Description	Impact
Define a standard data set for batteries	Given the large number of sensors embedded on a BESS, there is a need to be selective on the data to be logged. Data should be meaningful for advanced battery analytic algorithms.	Unified interfaces to the data space will enable different data provider/uses to share tools and knowledge for battery operation (such as battery analytic tools)
Ensuring privacy and security throughout the life cycle of the BESS	Share data, tools and analytic results through a centralised platform that assures privacy and security throughout the life cycle of the BESS.	Promote cooperation among different business operating in the area of BESS
Use the results of the data space to improve reliability of BESS	LIBATTION is aiming to take advantage of the results of the analytic tools developed within the use case 5 to improve the reliability of the BESS.	Increased reliability of BESS translates in greater customer satisfaction and greater remuneration
Access to CAD/Data from battery manufacturers (EV & ESS)	Gaining and streaming access to design data of batteries from like CAD, datasheet, etc.	Enables SIPBB to create highly accurate digital twins and generate synthetic data which in turn can be used to train AI models

Providing synthetic data, digital twins and trained models	Providing digital twins, synthetic data and trained AI models to battery operators and recyclers	Enables streamlined analysis and predictive maintenance for battery operators; and enables efficient automation of battery de-manufacturing or disassembly for recycling
Develop federated cross-domain analytics focused on accurate estimations of battery parameters of the BESS	Accurate predictions and calculations of batteries' status with the development and training of AI/ML models (Battery Analytics)	Improves the reliability to Libattion towards their customers and helps serve Green Deal objectives. The analytics will potentially facilitate energy efficiency by providing details about battery status. The batteries could potentially enter a circular economy ecosystem
Achieving energy efficiency in terms of digital resources	Develop AI models with cross-domain applications, able to be applied to different stakeholders in a federated way	Reduced digital resources, by reusing models instead of developing and using different models for the same purpose

3.6.5 Actors involved

Table 19 UC5 actors involved

Actor	Business area	Type of impact	Description of the impact
Customers	System Reliability & Profitability	Direct	The Battery Analytics Tools (developed by specialist analysts using the battery data on the cloud) will provide more accurate state of the system and predict more accurately its maintenance, reducing the system downtime and increase profitability. The customer (owner of the BESS) will also have direct access to visualise the state of the system.
Company managers	Management	Direct	Managers will gain better oversight of production efficiency and system performance, aiding in decision-making and process optimization.

3.6.6 User & system requirements

Table 20 UC5 requirements

Business objective	Requirement
Access to CAD/Data from Battery Manufacturers (EV & ESS)	Connector, IP-contract with Data-provider

Providing synthetic data, digital twins and trained models	Connector, source data (from above)
Develop federated cross-domain analytics focused on accurate estimations of battery parameters of the BESS	Access to a sufficient amount of BESS labelled data and domain knowledge
Achieving energy efficiency in terms of digital resources	Access to the integrated PLIADES framework for testing the cross-domain capabilities of data

3.6.7 Expected benefits and KPIs

The UC5 of the PLIADES project is expected to lay the foundation for a simplified platform to share data, knowledge and dedicated tools in the field of BESS. The platform will allow easier cooperation between battery manufactures and users, as well as suppliers and service providers. To do so, a certain level of standardization and ruling is required. The standardization and ruling of the platform might be a continuous process, and the first definition of it is done at this initial stage.

Table 21 UC5 benefits and KPIs

Indicator	Description	Current value	Future expected value	Expected date of achievement
Data Standardization	The standard data set is a document that defines what data should be logged for each battery.	None	Full definition	During the implementation
Battery Digital Twin	SIPBB to elaborate battery digital twin that will serve to improve the assembly/dismantling/maintenance processes.	None	Availability of digital twin	Before 6 months after implementation
Battery analytics on the cloud	Develop and implement battery algorithms capable to better estimate battery status and predict its maintenance using the data space data logs. Also, provide data visualization through interactive dashboards.	None	Availability of battery analytic tools	Before 12 months after implementation

3.7 Use Case 6: Integrating professional service robot data life cycles to improve HRI with robot operators

3.7.1 Use case overview

Assistive robots can enhance safety, efficiency, and productivity in various domains. This use case focuses on healthcare and industrial settings. In healthcare, robots can reduce strain on staff and improving patient care, especially as the ageing population drives demand for more support. In industrial environments, robots help with repetitive tasks, material handling, and quality control, reducing workplace accidents and improving efficiency. By working alongside human workers, they address labour shortages, improve safety, and support higher productivity. This use case focuses on improving the interaction between robots and robot operators in both domains. Professional service robots need to be endorsed with further advanced capabilities for smooth and efficient HRI to promote their adoption in real healthcare and industrial environments. Although deep-learning - based approaches gain more and more ground in this scope, they require vast amounts of relevant data to be collected and become available for re-use to achieve effective training. Moreover, these data refer to human commands from different modalities (e.g., speech, gestures) that have to be translated in specific robot plans and actions. Nevertheless, they might include different levels of abstraction or be affected by significant human factors with respect to individual and cultural specificities.

In this use case, HRI data of three different scenarios will be considered: i) interacting with a telepresence robot in a production facility, ii) operating the PTR Robot in a hallway, iii) nursing personnel support in patient monitoring, and iv) a robot that performs inspection in a manufacturing setting.

3.7.2 Background

3.7.2.1 Participants

Blue Ocean Robotics (BOR) develops, produces and sells professional service robots primarily in healthcare. The company develops and introduces new innovative market-ready robots for healthcare, pharmaceutical, hospitality, and other global markets. The company develops robots from problem, idea and design to development, commercialization, and all the way to exit. Each robot brand is set up in its own subsidiary- venture company focused on commercialization, global distribution and growth. BOR will act as a data user and data provider for the Healthcare and the Industrial Data Spaces.

The Centre for Research and Technology Hellas (**CERTH**) is a research organization working on various topics related to robotics and human-robot interaction (HRI), including the development of human action recognition methods, human-aware navigation, equipment and fault detection. The healthcare and manufacturing domains are settings this research has targeted through the years, and PLIADES will be an opportunity to further expand and improve the developed methods, while also producing valuable data for use in data spaces. CERTH will act as a data user and data provider for the Healthcare and the Industrial Data Spaces.

I4byDesign is a Competence Centre, focused on helping SMEs towards the adoption of Industry 4.0 & Logistics 4.0 technologies. Our vision is to combine know-how, experience, and technical expertise of its shareholders on Industry 4.0 technologies, in order to promote the application of integrated technological solutions. Our mission is to contribute, in National and International level, to the acceleration of digital and technological transformation of the Greek industrial ecosystem, focused on SMEs. I4BD will assist UC6c, by providing additional equipment for the robots and a testbed for data collection.

3.7.2.2 Data spaces readiness

BOR has not worked with Data Spaces before joining the PLIADES project and thus sees the project as business opportunity to improve its robots with the use of data spaces.

CERTH has no prior experience on the usage of data spaces, but in contrast has great experience in the development of AI methods for mobility applications.

I4BD has no prior experience with data spaces. I4BD views data spaces as a strategic enabler for secure, interoperable data sharing, aligning with its mission to deliver cutting-edge solutions. By participating in PLIADES, I4BD aims to enhance the expertise of its personnel, build strategic collaborations with key players from around EU and expand its business opportunities by providing competitive and innovative solutions.

3.7.2.3Current business scenario

3.7.2.3.1 HRI in telepresence robot operation

The GoBe Robot (Figure 12) can be used for site inspections in a manufacturing context. The robot enables the site inspector to move through buildings, perform visual assessments, interact with locals, and more. The robot is controlled remotely via a programme on the PC by the person on the telepresence robot. Currently, no data is collected by the robot and shared in a data space. However, the robot is equipped with hardware to enable it to collect data.

Figure 12 GoBe Robot in a production setting today

3.7.2.3.2 HRI in rehabilitation robot operation

The robot needs to be transported to different locations inside the hospital. Currently, the robot is operated by a human (see Figure 13). It is envisioned that the robot can either follow the operator or drive autonomously to locations. This would make the transport more pleasant and efficient for the operator. Currently, no data is collected by the robot and shared in a data space. The robot is also not equipped with specific hardware to collect data.

Figure 13 Operator driving robot via joystick to destination

3.7.2.3.3 HRI in nursing personnel support

Socially assistive robots can be valuable in supporting the nursing personnel in a hospital setting. The focus lies in improving the interaction between the caregivers and the robot, when an assistive mobile robot is used for patient monitoring. Accurate human action recognition AI-based methods, human-aware navigation techniques and human-robot interaction schemes have been utilized by CERTH to that end in the healthcare setting. Various sensors have been employed with occasional data collection, to train and improve the models.

Currently, on any CERTH robot that performs these tasks, the data is processed and dropped during normal (live) operation without utilising it for later use. In cases where the captured data should be stored or exploited, a process is explicitly triggered for that purpose. This is not yet an established process but could become a realistic service in the future with the help of PLIADES technologies.

3.7.2.3.4 HRI in manufacturing inspection robot operation

Utilizing autonomous mobile robots for equipment inspection and for ensuring the safety of the workers in a manufacturing setting is proving to be efficient and evermore important. To that end, these two scenarios will be explored and improved in this use case with a focus on the interaction between the robot and its operator. Existing methods for human action recognition, equipment inspection, human-aware navigation and HRI will be employed and expanded upon.

Currently, on any CERTH robot that performs these tasks (see Figure 14), the data is processed and dropped during normal (live) operation without utilising it for later use. In cases where the captured data should be stored or exploited, a process is explicitly triggered for that purpose. Evidently, this is not yet an established process but could become a realistic service in the future with the help of PLIADES technologies.

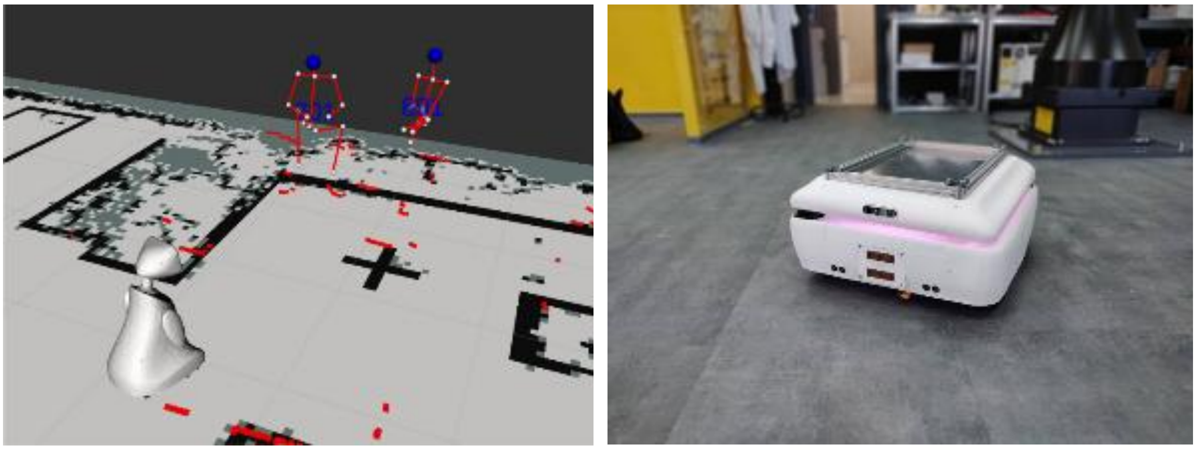

Figure 14 Robots from CERTH performing human action recognition & inspection tasks

3.7.2.3.5 Current Data Processes

The data processes that are common across the involved partners of this scenario are the following:

- Data collection
- Data filtering
- Data synchronization
- Data annotation
- Data sharing

Data collection refers to the production and storage of raw multimodal sensor streams from autonomous robots and their input devices. This process starts on demand and stores the data locally or streams the sensor measurements and completes with the storage of these data into a server. This currently leads to mass amount of data, making it challenging to select the most useful parts of the collected data with the optimal quality that will assist the AI development.

During the data creation step, the format of the generated data may not be the same among the various robots and partners. Even though BOR and CERTH utilize ROS, the created dataset might not be created in the same way, thus leading to incompatibilities between the partners.

These data are saved in local servers of the data provider without considering any sustainable practices, leading to massive incoming data and potential burden when shared.

Data filtering refers to the execution of processes that improve the data quality by removing duplicates, excluding low-quality or noisy frames (e.g., blurred images, silent chunks of audio), and removing corrupted frames and generally eliminating irrelevant information that could lead to potential confusion in the system. Currently, data filtering is performed offline as part of the data pre-processing steps. This process involves manual reviewing of the collected data by human experts, followed by manual data exclusion in cases where it is needed.

Data synchronization refers to the process of establishing consistency between the data originating either from different sensors on the same agents or from different agents. So, robots equipped with multiple sensors are heavily reliant on the accurate synchronization of the sensor streams, so that all of them can be used for processing with accurate measurements. Existing tools, such as message filters provided for ROS-based infrastructures provide such functionalities. Since robots from BOR and CERTH use ROS, the only issue will be to ensure that the related stamps are always saved along with the data.

Data annotation refers to the process of generating labels for the collected data. These labels might refer to various AI-based tasks, such as object detection and semantic segmentation, or might be the recognition of events. The format of the labels can be either visual (image masks), text description or even audio labels. Also, the annotation can be related to different types of data and can have different formats for each one, e.g., the labels for point cloud data can be different that the label for image data. Currently, the data annotation process is mostly performed manually using some publicly available tools, either online or offline to extract the corresponding labels, or using some semi-automatic AI-based labelling pipelines that are executed in the cloud; however, they always need human verification. Another aspect of the data annotation process is the format of the labels. Currently, there is not standard followed by all partners and there are several available formats that are used by AI-based methods, such as YOLO format, COCO format etc.

Data sharing refers to the data interchange among various partners who share common needs and would benefit from having access to data from different conditions, sensors or infrastructure.

Currently, data sharing is not an automated process and requires great effort to adapt the available datasets, so that they will be usable. From the perspective of data providers, existing data sharing practice is to upload the generated datasets on a publicly available server, so that everyone has access to the data or to private servers, where access is selectively provided. In both approaches, this process requires direct contact between the two parties (provider and consumer).

On the side of data consumers, the data are downloaded locally to a computing unit, their integrity cannot be guaranteed, while improvement by the data provider cannot be performed online but requires complete transfer of the dataset. Also, the data consumer needs to manually search for the available datasets, through online catalogs and select the optimal dataset, based on samples provided and their description.

3.7.2.4Pain points

Social and Ethical Aspects: Ensuring the safety of individuals passing near the robot during its trials is a critical concern. This raises ethical questions about how to protect people while testing the robot in real-world environments. The robot must also be designed to operate in a way that respects societal norms and privacy, especially during data collection activities.

Technical Aspects: The robot must be capable of determining whether a situation is challenging enough to require external assistance. Capturing data dynamically while ensuring compliance with privacy-preserving practices and GDPR regulations is another significant challenge. Identifying and recognizing the trigger points that signal the need to begin automatic data collection and storage is essential. Additionally, processes and communication channels must be established for transferring data from the robot to designated Data Spaces. Managing the substantial amount of data that needs to be stored presents a further technical challenge. If data is stored locally, disk space could become an issue, while streaming it externally may lead to bandwidth and networking constraints.

3.7.2.5Barriers

3.7.2.5.1 Legal framework

The EU's legal framework for assistive robots in manufacturing, focusing on mobile robots, is primarily governed by safety directives (Machinery Directive, General Product Safety Directive), ethical guidelines, and standards (such as ISO 10218 and ISO/IEC 13482). The EU’s legal framework for healthcare robotics and AI combines medical device regulations, data protection laws, liability standards, and ethical guidelines. Legal frameworks, apart from the ones described in Section 3.1.1, focusing on the manufacturing domain are:

- EU Machinery Directive (2006/42/EC): This directive sets out the essential health and safety requirements for machinery, including industrial robots and assistive robots used in manufacturing. The directive ensures that robots meet safety standards to minimize risks to workers and operators. It covers aspects such as design, construction, installation, and maintenance of machines. It is important that mobile robots in manufacturing are CE marked, indicating compliance with this directive.

- Standards (ISO 10218/ISO/IEC 13482:2014): These international standards for industrial robots, adopted in the EU, focus on the safety of industrial robots. They cover the design, installation, operation, and maintenance of robots to ensure safe interaction with human workers and the environment. The standards are relevant for collaborative robots (cobots) that work alongside human workers in manufacturing settings.

Legal frameworks focusing on the healthcare domain:

- Medical Device Regulation (MDR): In the EU, robotics and AI used in healthcare often fall under medical device regulations. The MDR (EU 2017/745) ensures that medical devices, including robotic systems and AI applications, are safe and effective. These regulations require conformity assessment before market entry and ongoing post-market surveillance.

3.7.2.5.2 Social framework

The adoption of robots in manufacturing and healthcare holds immense potential to improve efficiency, safety, and quality of life. Assistive robots in manufacturing are designed to optimize productivity, safety, and efficiency. Despite these benefits, the widespread adoption of robots in manufacturing has fuelled fears of job displacement. Workers often worry that automation will render their roles obsolete, leading to unemployment and economic insecurity. Fears of job displacement can exacerbate economic insecurity, highlighting the need for policies that support workforce re-educate and job creation in complementary roles.

In healthcare, robots are supporting patient care, surgery, and elder support. Robots designed for elder-care can assist with daily activities, monitor health, and even offer companionship to ageing populations. Assistive robots in healthcare are designed to help with daily activities allowing individuals with disabilities or elderly people to maintain independence and improve their quality of life. However, the introduction of robots into healthcare brings challenges related to trust, ethics, and human interaction. Patients and caregivers may feel uneasy about relying on machines for critical tasks, fearing potential malfunctions or a lack of empathy. Building trust requires developing robots that are safe, reliable, and capable of demonstrating culturally and emotionally aware behaviour.

PLIADES is addressing this issue in the use cases by enabling the robots to interact on a higher level with end-users, through data sharing with other robot manufacturers. By focusing on transparency and the development of culturally aware robots, society can ensure that these technological advancements benefit all stakeholders while fostering trust and acceptance.

3.7.2.5.3 Economic framework

The implementation of robots, both in manufacturing and healthcare, often requires restructuring workflows, which may lead to temporary losses in productivity and increased expenses during the transition period. Retraining programs are also required to help workers adapt to new roles, further adding to the financial burden. In healthcare specifically, the high cost of acquiring and maintaining robotic systems is a significant barrier for many institutions. Beyond the initial investment, robots demand ongoing expenditures for staff training, system updates, and maintenance. Costs can be reduced if robots become smarter and implement easier into organizations through data sharing. By making use of shared datasets as this use case suggests, robots can learn and improve their performance more rapidly, reducing the need for extensive individual programming. This enhanced intelligence can streamline implementation, making it easier to integrate robots into existing workflows while minimizing disruption.

3.7.3 Future scenario

3.7.3.1HRI in telepresence robot operation

A robot operator is currently being introduced to the production site and workflow of Blue Ocean Robotics in Denmark to familiarize themselves with the company and its processes. To facilitate this, another colleague in Denmark instructs the robot:

- by follow-me function
- to turn right or left

The pilot can also take over, which the robot indicates to the operator.

The following specific processes are included in the use case:

- The colleague from the US, who is the pilot (controlling the robot), is logging into the system at his PC and into the robot at the HQ in Denmark.
- At the HQ, his colleague is meeting the pilot on the robot at the docking station.

The robot is always at a docking station. The colleague in Denmark is picking the pilot up at the docking station to show him around. Normally, the pilot needs to navigate the robot to the place at the production site. To save time and for the pilot to focus fully on the production site tour, the colleague in Denmark is initiating the “follow-me” function. With specific gestures, the Danish colleague can now instruct the robot to:

- Follow the person on-site: the robot follows the person at the HQ, this makes it easier and faster to move the robot and the pilot can focus on e.g., checking out the surroundings at the production site. HRI: The person interacts with the robot via gestures to indicate the robot to follow or stop. The robot communicates with light or sound.
- Turn the robot right or left in a specific direction: The person can also direct the robot’s screen/camera in certain directions with pre-defined gestures to e.g., turn the robot towards a specific item that the pilot should check out (the camera through which the pilot is watching is above the robot screen, the pilot can only see the area in which the robot screen is directed to - the robot does not have a camera on the back, right or left). HRI: The person interacts with the robot via gestures to indicate the robot to turn right or left. The robot communicates with light or sound.
- The pilot takes over if s/he wants to turn the robot into a specific direction. This should be indicated on the robot, so the person on site knows about it.
- The GUI is occupied with the video stream of the colleague in the US, it can still be used for communication, but speech commands might be easier.

3.7.3.2 HRI in rehabilitation robot operation

The process for this scenario is that the robot is needed at a specific location for a rehabilitation exercise. The robot navigates from one location (charging station) to another (rehabilitation room) via follow-me function by a healthcare professional. Commands by human are given via gestures, robot confirms via light/sound.

The specific steps in this process are as follows:

- The robot is picked up by the staff and follow-me navigation is initiated by a gesture, to navigate through the corridor.
- The robot follows the staff, the staff is confirming the follow me function by gestures until the location is reached, there the staff is indicating stop with a gesture.
- The robot can provide feedback to staff in case it experiences difficulties by sound, light, GUI.
- The staff can switch to joystick to help robot navigate e.g., around obstacles or through a door.

- Once arrived at the location, the robot is used for the rehabilitation session.

3.7.3.3 HRI in nursing personnel support

The patient care process begins when nursing personnel assign a care or monitoring task to the robot using its GUI or verbal commands. The caregiver provides key details, including the patients' location, and instructions, such as monitoring duration.

Once the task is assigned, the robot confirms receipt and offers an opportunity for the nurse to verify or adjust the task parameters. Afterward, the robot navigates autonomously to the designated locations, providing real-time updates to the nursing staff.

During the monitoring period, the robot communicates directly with the nursing staff through a dedicated GUI, providing real-time updates on the detected actions of the patient. Caregivers can also send commands to the robot remotely, such repositioning the robot for a better vantage point.

Upon completion of the task, the robot provides a report summarizing the monitoring period. This report is transmitted to the nursing personnel through the GUI and can also be stored for record-keeping. The robot then returns to its initial state and position, ready for further instructions from the nursing team.

3.7.3.4 HRI in manufacturing inspection robot operation

Operator/worker safety inspection scenario: The safety inspection process begins with the robot being assigned a specific area of the building to patrol. This assignment, including information about the designated area, is provided by the user through the robot's graphical user interface (GUI).

Upon receiving its instructions, the robot autonomously navigates to the specified location. Once there, the robot begins patrolling the area, focusing on workers and their surroundings to identify and alert for potential safety hazards. These hazards might include situations such as a worker being too close to dangerous machinery or exhibiting poor body posture that could lead to injury.

If the robot detects an unsafe situation, it approaches the worker involved and issues an alert via a voice message. To ensure the worker has received and understood the message, the robot waits for feedback, which can be provided through voice, gesture, or by using the robot's GUI.

After verifying that the worker is aware of the situation, the robot leaves the immediate area and resumes its patrolling and inspection duties.

Equipment inspection: An inspection task begins when an operator or worker assigns the robot to inspect a specific area of the building via the robot's graphical user interface (GUI). The operator provides details about the area to be inspected through this interface.

Upon receiving its instructions, the robot autonomously navigates to the designated location. Once there, the robot performs the inspection, actively identifying malfunctioning equipment and potential issues within the infrastructure.

If the robot detects a problem, it returns to the operator or the person responsible for building maintenance and reports its findings. Following this report, the robot offers the operator the option of a "follow-me" scenario, guiding them directly to the precise location where the issue was detected.

In both cases, the technologies developed in PLIADES will facilitate the data process, enabling continuous and smart data collection from any modality of the robot that relates to the interaction with the operator, such as GUI interactions, voice commands, gestures etc. These in turn will be utilized to enhance any models used for HRI.

3.7.3.5 Future scenario data processes

The processes that will be affected by the PLIADES framework are relevant to all business scenarios that belong to use case 6. The business processes are directly related to the data operations, which were mentioned in the previous section, and that are performed either by data providers or data consumers or both.

Data collection's evolution is related to the standardization of data formats that are used by the involved partners, as well as to the introduction of sustainable practices, leading to the green approaches for data collection.

The format of the collected data will be based on standardized formats used in robotics and HRI. Such formats are common file formats that are widely used for images, sound, videos, etc. such as .JPG, .MP4, .PNG, etc. Also, data collection will use a common format based on XML structures to describe the relative placement among different sensors and also any other necessary metadata that can assist the time and location synchronization of different data sources.

Sustainable practices will be established in order to reduce the amount of data collected, but without compromising the information available. These AI-based methods will be executed either real-time, during the streaming of sensor data to a computing unit, or offline, as a post-processing step to the collected data.

Through the PLIADES framework, **data filtering** will undergo important improvements that will lead to improved performance of these processes, while also reducing the required time. Specifically, data filtering will be performed online, on the edge, by performing a sequence of filtering pipelines in a specified order. The outcome generated will be an updated version of the data with increased quality, along with quality metrics for various types of data.

These pipelines will evaluate the data quality and will automatically discard data that do not contribute to the collected data sets. AI-based approaches will be developed as part of the overall data processing pipeline and will consider data quality aspects to perform automated data filtering. In addition, specific modules for privacy-related filtering will be developed, in order to either discard frames or edit them accordingly, in order to follow the privacy guidelines. Such methods include the blurring of faces in image data and other potential needs that will be extracted from the related tasks.

Regarding **data annotation**, the existing automated annotation methods will be further expanded with human knowledge injection methods.

Also, annotation will be based on standardized formats. An appropriate ontology might also be used to match the data with real-world entities. Data annotation will also include the handling of metadata insertion, suitable for the improved discoverability of data. To this end, automated annotation tools will be developed.

Data sharing will undergo significant improvements by the incorporation of data spaces and the PLIADES framework. The previous data processes will prepare the data in a way that will allow seamless exchange among partners, allowing them to directly use them.

From the perspective of data providers, the data sharing process uses as input the data that have been generated by the previous processes of Collection, Filtering, Synchronization and Annotation. These data will be extended with semantics metadata that will contain all the essential information regarding the ownership of data, the access rights and the potential regulations of usage. These data with their metadata will be available to the AI-based brokering schemes and provided through the developed AI-based connectors. Data providers will also be able to actively maintain and update their provided datasets.

From the perspective of data consumers, they will first of all make use of the AI-based querying system and the AI-based brokers in order to identify all the available datasets within multiple dataspaces and receive an accurate suggestion on a dataset, according to their needs. Data consumers will also agree on the pre-defined policies for the data usage, as part of their connection on the dataspace, ensuring the proper use of data.

3.7.3.6Cross data space connections in the use case

BOR will act as **data provider** and **data consumer** in the **healthcare and industrial** data space. Specifically, BOR will provide data from the PTR Robot and the GoBe Robot (LiDAR, RGB images). BOR will also consume data related to HRI from the healthcare data space and the industrial data space.

CERTH will act as **data provider** and **data consumer** in the **healthcare and industrial** data space. CERTH will also consume the same types of data that might originate from other data providers, in order to develop explainable AI models for (re-)training for human robot interaction in the healthcare and industrial data spaces.

3.7.4 Objectives

Table 22 UC 6 business objectives

Business objective	Description	Impact
Enable seamless HRI by leveraging advanced Deep Learning methods to gather and analyse operational data from assistive service robots in healthcare and manufacturing	This will allow robots to continuously learn and adapt to the unique needs and behaviours of individual users. By e.g., enabling real-time data collection during robot operation, PLIADES will enhance the robots' ability to personalize interactions, improve rehabilitation outcomes, and optimize patient or eldercare support over time.	Achieving this objective will improve the following: – Trust and acceptability of robot operator – Economic benefits of robotic implementation, contributing to a more efficient, cost-effective, and sustainable healthcare system.
Ensure privacy, security, and traceability of data	With the introduction of data collection and data spaces, involved stakeholders will likely worry about privacy and security of the collected data, specifically if it is patient related. Thus, it will be necessary to implement robust measures to protect the confidentiality and integrity of the collected data, ensuring compliance with regulations and maintaining detailed records of data provenance and usage.	Achieving this objective will support the trust and acceptability of the stakeholders.

3.7.5 Actors involved

Table 23 UC6 actors involved

Actor	Business area	Type of impact	Description of the impact

Healthcare provider - management level	Management	Direct	The management level at the healthcare provider will evaluate the technical results and assess (according to economic aspects, security, acceptability, etc.) whether it could be deployed at their facility in the future.
Healthcare professionals (physiotherapist, nurse, doctor)	Healthcare	Direct	Healthcare professionals will directly interact with the robot, potentially also oversee patient data and suggestions on possible healthcare programmes by the AI tools.
Patient	Others	Direct	Patients will use and interact with the robot that is collecting the data and making us of the data from the data space.
Robot manufacturer	Manufacturing	Direct	The robot manufacturer is providing the data and using data from the Data Space to optimize the robot's ability to interact with humans.
Data integration and analysis specialists	IT	Direct	Data integration and analysis specialists will be responsible for integrating, analysing, and preparing new data sources. They will benefit from improved data quality, accessibility, and more efficient research workflows, enabling them to conduct high-quality and accurate research.
IT security specialists	IT	Direct	IT security specialists will be responsible for implementing and maintaining enhanced security and privacy measures. This will include managing encryption, access controls, and monitoring systems to ensure data integrity and compliance with regulations.
Research project managers	Management	Direct	Research project managers will oversee the integration of new data sources, the implementation of data sharing protocols, and coordination between departments and external partners. They ensure smooth data flows, adherence to timelines, and alignment with research objectives.
Compliance officers	Legal	Direct	Compliance officers will ensure that all data handling and sharing practices comply with relevant regulations and standards. They will be involved in auditing data processes and maintaining compliance documentation, and ensuring adherence to privacy laws

Financial administrators	Accounting	Indirect	Financial administrators will manage the financial aspects of the data integration and sharing initiatives. They will handle budgeting, funding allocation, and financial reporting related to these projects, ensuring proper financial management and accountability.

3.7.6 User & system requirements

Table 24 UC6 requirements

Business objective	Requirement
Improved data quality and annotation consistency to enable seamless HRI by utilizing advanced ML/DL methods	Implement compression techniques for data transmission
	Optimize edge computing for on-board filtering processes
	Automated identification and discard of noisy and low-quality multimodal data
	Semi-Automated methods for data quality estimation for multimodal data
	Automated metadata tagging during data collection
	Increased discoverability of data through context-aware metadata
	Automated analysis for cultural issues of human tracking data
Sustainable and cost-efficient data operations	Definition of policies for sustainable data operations
	Automated annotation processes for multimodal data
	Implement compression and filtering techniques to reduce data storage needs
	Automated identification and discard of redundant data
Ensure privacy, security, and traceability of data	Implement advanced encryption protocols for all data
	Automated monitoring of data adherence to privacy and quality standards
	Implement data anonymization techniques to protect sensitive information
	Set up a system for tracking data usage, sharing metrics, and maintaining version histories to ensure data traceability

3.7.7 Expected benefits and KPIs

The expected benefits of UC6, which focuses on integrating professional service robot data lifecycles to improve Human-Robot Interaction (HRI) with operators, include enhanced operator efficiency by streamlining interaction workflows and minimizing the need for manual control. By leveraging operational data, the system facilitates continuous improvements in HRI capabilities, ensuring robots operate with greater precision and adaptability. Real-time feedback mechanisms improve safety and reliability, making robot operations more robust. The integration of diverse data sources through data spaces enables scalability and interoperability, fostering collaboration across different sectors. Additionally, the adoption of greener data processes contributes to sustainability by reducing energy consumption and operational costs, aligning with broader environmental goals and advancing innovation in professional service robotics.

Table 25 UC6 Benefits and KPIs

Indicator	Description	Current value	Future expected value	Expected date of achievement
Increased usability	The robot will support robot operators when controlling the robot, resulting in improved usability. The System Usability Score (SUS) is a well-known method to evaluate usability of a product. For the GoBe Robot, there is currently no SUS data yet. It is expected to achieve similar results as the improved PTR Robot.	69/100	80/100	Once technology is at TRL9
Increased efficiency of robot installation and integration in daily workflow	With a more autonomous and interactive robot, installation at the healthcare or production site will be more efficient, especially if a robot is introduced to a new department internally or new personnel needs to become familiar with the robot. One aspect where the PTR Robot scored low on the SUS is “learn to use it quickly”. We expect this to increase after PLIADES. For the GoBe Robot, there is currently no SUS data yet. It is expected to achieve similar results as the improved PTR Robot.	3.5/5	4.5/5	Once technology is at TRL9
Increased trust and acceptability of robot operator in the robot	The robots’ ability to personalize interactions is expected to increase trust and acceptability of robot operators. One aspect where the PTR Robot scored low on the SUS was “confidence in operating the robot”. We expect this to increase after PLIADES. For the GoBe Robot, there is currently no SUS data yet. It is	3/5	4/5	Once technology is at TRL9

	expected to achieve similar results as the improved PTR Robot.			

3.8 Definition of PLIADES use cases

The following table summarizes the high-level use cases and their sub-use case scenarios.

Table 26 Summary of high-level use cases and their sub-use cases

High-level use case	Sub-use case ID and title
UC1. Integrating data life cycles of sustainability, operations and process industry manufacturing operations (Industrial Data Space)	SubUC1.1. Increment DENN data platform penetration in reluctant sectors through data compliance and security
	SubUC1.2. Data-Driven Lifecycle Management for Suppliers
	SubUC1.3. Commercial Expansion of DENNDATA for Cross-Industry Data-Driven Optimization
UC2. Integrating data life cycles of service robot to improve HRI with end users (Healthcare Data Space)	SubUC2.1. HRI in rehabilitation
	SubUC2.2. HRI in patient monitoring
UC3. Integrating data life cycles of personalized medicine services to improve diagnostic and prognostic clinical prediction models (Healthcare Data Space)	SubUC3.1. Sharing Data and Models among Partners
	SubUC3.2. Collaborative Data/Service Exchange and Third-Party Data Acquisition
UC4. Integrating data life cycles of smart vehicles for CCAM operations and ADAS/AD functions (Mobility and Industrial Data Spaces)	SubUC4.1. AI-based ADAS development
	SubUC4.2. AI-based traffic management
UC5. Integrating data life cycles of WEEE/batteries management and car parts manufacturing operations (Green Deal and Industrial Data Spaces)	SubUC5.1. Sharing field data from battery systems energy storage installations
UC6. Integrating professional service robot data life cycles to improve HRI with robot operators (Healthcare and Industrial Data Spaces)	SubUC6.1. HRI in telepresence robot operation
	SubUC6.2. HRI in rehabilitation robot operation
	SubUC6.3. HRI in manufacturing inspection robot operation

4 Requirements and analysis

The following section contains all collected user and system requirements which have been gathered from the different use cases through the Business Requirements Questionnaire. A total of 90 requirements have been gathered for the 6 PLIADES use cases as can be seen in Table 27 for the functional requirements and in Table 28 for the non-functional requirements. Following the previous analysis of the use cases that ended up with the identification of the requirements, because of their similarities Use Case 2 and Use Case 6 have aligned that identification so those requirements that have been identified for use case 2 are also valid for use case 6, and so, any further analysis will take into account this in terms of numbers and relevance.

Table 27 Functional user requirements of the PLIADES platform

ID	Type	Description	Priority Level
UR1	Functional	Implement automated quality checks based on real-time machine data	HIGH
UR2	Functional	Enable flexible machine settings adjustments based on varying production requirements	MEDIUM
UR3	Functional	Provide adaptive control for material specifications based on data-driven insights	HIGH
UR4	Functional	Integrate quality data from previous production runs to improve current batch consistency	MEDIUM
UR5	Functional	Facilitate collaborative adjustments between operators and quality managers for improved outcomes	LOW
UR6	Functional	Use predictive algorithms to adjust machine parameters before defects occur	HIGH
UR7	Functional	Enhance data visualization tools for real-time monitoring of quality metrics	MEDIUM
UR8	Functional	Dynamically adjust production schedules to account for material and product specification changes	LOW
UR9	Functional	Integrate human-in-the-loop methodologies to refine AI algorithms	HIGH
UR10	Functional	Automate compliance checks for data access requests	MEDIUM
UR11	Functional	Provide users with real-time compliance notifications for data operations	HIGH
UR12	Functional	Implement pilot projects with early adopters in new markets	HIGH
UR13	Functional	Develop training programs to help clients utilize DENNDATA for new business opportunities	HIGH

UR14	Functional	Leverage AI to explore cross-industry data opportunities	LOW
UR15	Functional	Enable real-time collaboration tools for stakeholders working on new projects	LOW
UR16	Functional	Implement compression techniques to optimize data transmission	HIGH
UR17	Functional	Optimize edge computing for onboard filtering processes	MEDIUM
UR18	Functional	Automate identification and discard of noisy and low-quality multimodal data	HIGH
UR19	Functional	Develop semi-Automated methods for multimodal data quality estimation	MEDIUM
UR20	Functional	Automate metadata tagging during data collection	HIGH
UR21	Functional	Increase data discoverability through context-aware metadata	MEDIUM
UR22	Functional	Automate analysis for cultural issues of human tracking data	LOW
UR23	Functional	Automate annotation processes for multimodal data	HIGH
UR24	Functional	Implement compression and filtering techniques to reduce data storage needs	MEDIUM
UR25	Functional	Automate identification and discard of redundant data	HIGH
UR26	Functional	Implement advanced encryption protocols for all data	HIGH
UR27	Functional	Automate monitoring of data compliance with privacy and quality standards	MEDIUM
UR28	Functional	Implement data anonymization techniques to protect sensitive information	HIGH
UR29	Functional	Develop a system for tracking data usage, sharing metrics, and maintaining version histories to ensure data traceability.	HIGH
UR30	Functional	Develop a centralized data repository for storing integrated health research data and implement data integration workflows to automate data ingestion from various sources	HIGH
UR31	Functional	Implement APIs for seamless data sharing with external partners	HIGH
UR32	Functional	Develop user-friendly data visualization tools	MEDIUM
UR33	Functional	Develop a comprehensive access control system to manage user permissions	HIGH
UR34	Functional	Define and adopt standardized formats for annotation labels	HIGH

UR35	Functional	Automate conversion of existing datasets format into standardized formats	MEDIUM
UR36	Functional	Enable seamless data exchange between systems and devices through easy-to-use APIs	MEDIUM
UR37	Functional	Validate incoming data against predefined schemas and standards	MEDIUM
UR38	Functional	Create a metadata repository for standardized formats and protocols	MEDIUM
UR39	Functional	Optimize edge computing for on-device data processing and filtering, eliminating the reliance on centralized systems	MEDIUM
UR40	Functional	Automate deletion of outdated and expired data	MEDIUM
UR41	Functional	Allow real-time processing of multimodal data	MEDIUM
UR42	Functional	Integrate AI-driven insights and analytics into resource planning	HIGH
UR43	Functional	Conduct real-time quality assessment of multimodal data	MEDIUM
UR44	Functional	Provide intuitive interface for semi-automated annotation processes	HIGH
UR45	Functional	Inject human knowledge during the annotation processes	HIGH
UR46	Functional	Provide feedback to annotator for refining AI-generated annotations	MEDIUM
UR47	Functional	Develop tools for real-time visualization of annotations	HIGH
UR48	Functional	Generate reports on the annotation quality and consistency of data	MEDIUM
UR49	Functional	Support metadata tagging in annotations to include context-specific information	HIGH
UR50	Functional	Provide secure data-sharing infrastructure with encryption mechanisms	HIGH
UR51	Functional	Enable datasets discoverability with AI-based brokering mechanisms and recommendation engines	HIGH
UR52	Functional	Implement role-based access control for data access	HIGH
UR53	Functional	Provide real-time anonymization of sensitive and personal data	HIGH
UR54	Functional	Define a standard data set of parameters to be logged for lifetime monitoring of BESS	MEDIUM

UR55	Functional	Develop and deploy necessary tools for collecting data based on defined standard dataset	HIGH
UR56	Functional	Share necessary BESS data via the dedicated PLIADES Data Space (e.g., lifetime monitoring data, CAD, datasheets, Bill of Materials, etc.)	HIGH
UR57	Functional	Manage data accessibility for different stakeholders within the platform	HIGH
UR58	Functional	Develop dashboards for visualizing battery data life cycles	HIGH
UR59	Functional	Develop analytics tools for life cycle data (e.g. battery state estimation and BESS predictive maintenance)	HIGH
UR60	Functional	Create accurate predictions of battery status with the development and training of AI/ML models	HIGH
UR61	Functional	Develop cross-domain AI models for federated applications	HIGH

Table 28 Non-functional user requirements of the PLIADES platform

ID	Type	Description	Priority Level
UR62	Non-Functional	Utilize historical data to anticipate production bottlenecks and adjust processes proactively	MEDIUM
UR63	Non-Functional	Generate alerts for deviations in material quality based on predefined thresholds	HIGH
UR64	Non-Functional	Establish data governance policies to ensure data quality and consistency	HIGH
UR65	Non-Functional	Enable secure data sharing protocols across stakeholders, maintaining data sovereignty	MEDIUM
UR66	Non-Functional	Develop data anonymization techniques for sharing insights without compromising sensitive information	HIGH
UR67	Non-Functional	Develop tools for tracking data usage and auditing trails for accountability	HIGH
UR68	Non-Functional	Use advanced encryption techniques for data stored on cloud servers	MEDIUM
UR69	Non-Functional	Establish a framework for continuous data quality monitoring and improvement	MEDIUM
UR70	Non-Functional	Enable data-driven insights for regulatory reporting	LOW

UR71	Non-Functional	Develop partnerships with technology companies to integrate DENNDATA into their solutions	HIGH
UR72	Non-Functional	Identify potential new sectors for DENNDATA expansion, such as healthcare or energy	MEDIUM
UR73	Non-Functional	Establish co-innovation programs with clients to develop new use cases for DENNDATA	MEDIUM
UR74	Non-Functional	Use predictive analytics to identify emerging market trends	MEDIUM
UR75	Non-Functional	Automate the generation of market reports based on data insights	HIGH
UR76	Non-Functional	Establish legal frameworks for co-development agreements with clients and partners	HIGH
UR77	Non-Functional	Define policies for sustainable data operations and archiving	MEDIUM
UR78	Non-Functional	Provide training programs on new data sharing protocols and tools	MEDIUM
UR79	Non-Functional	Develop training programs on data privacy and security protocols	MEDIUM
UR80	Non-Functional	Define and adopt standardized data formats	MEDIUM
UR81	Non-Functional	Define and adopt standardized metadata formats	MEDIUM
UR82	Non-Functional	Monitor resource usage to optimize data operations	MEDIUM
UR83	Non-Functional	Ensure effective and scalable data operations	MEDIUM
UR84	Non-Functional	Implement cloud-based data operations for cost effectiveness	MEDIUM
UR85	Non-Functional	Train annotators on the use of the automated tools	HIGH
UR86	Non-Functional	Define and establish numerical metrics to evaluate annotation quality and consistency	HIGH
UR87	Non-Functional	Ensure compliance with privacy regulations in data-sharing activities	HIGH
UR88	Non-Functional	Define data access policies to control data sharing	HIGH
UR89	Non-Functional	Manage data sharing agreements and permissions through user-friendly interfaces	HIGH
UR90	Non-Functional	Continuously monitor and log data-sharing activity	MEDIUM

In a first approach, while analysing the functional requirements of the 6 use cases of the PLIADES project by type of requirement the following classification was done:

- Technical
- Performance
- Reporting
- User
- Privacy
- Security
- Infrastructure
- Access
- Management
- Collaboration

In the following figures, a categorization by type for each functional requirement is presented:

Use Case	UR Nº	Description	Technical	Performance	Reporting	User	Privacy	Security	Infrastructure	Access	Management	Collaboration
Use Case 1	UR1	Implement automated quality checks based on real-time machine data	■									
	UR2	Enable flexible machine settings adjustments based on varying production requirements		■								
	UR3	Provide adaptive control for material specifications based on data-driven insights				■						
	UR4	Integrate quality data from previous production runs to improve current batch consistency			■							
	UR5	Facilitate collaborative adjustments between operators and quality managers for improved outcomes				■						
	UR6	Use predictive algorithms to adjust machine parameters before defects occur	■									
	UR7	Enhance data visualization tools for real-time monitoring of quality metrics	■									
	UR8	Dynamically adjust production schedules to account for material and product specification changes				■						
	UR9	Integrate human-in-the-loop methodologies to refine AI algorithms			■							
	UR10	Automate compliance checks for data access requests			■						■	
	UR11	Provide users with real-time compliance notifications for data operations				■						
	UR12	Implement pilot projects with early adopters in new markets				■						
	UR13	Develop training programs to help clients utilize DENNDATA for new business opportunities				■						
	UR14	Leverage AI to explore cross-industry data opportunities	■									
	UR15	Enable real-time collaboration tools for stakeholders working on new projects				■						
		Use Case 1 TOTAL	**4**	**1**	**3**	**7**	**0**	**0**	**0**	**0**	**1**	**0**

Figure 15 Use Case 1 functional requirements categorized by type

Use Case	UR Nº	Description	Technical	Performance	Reporting	User	Privacy	Security	Infrastructure	Access	Management	Collaboration
Use Case 2 -6	UR16	Implement compression techniques to optimize data transmission										
	UR17	Optimize edge computing for onboard filtering processes										
	UR18	Automate identification and discard of noisy and low-quality multimodal data										
	UR19	Develop semi-Automated methods for multimodal data quality estimation										
	UR20	Automate metadata tagging during data collection										
	UR21	Increase data discoverability through context-aware metadata										
	UR22	Automate analysis for cultural issues of human tracking data										
	UR23	Automate annotation processes for multimodal data										
	UR24	Implement compression and filtering techniques to reduce data storage needs										
	UR25	Automate identification and discard of redundant data										
	UR26	Implement advanced encryption protocols for all data										
	UR27	Automate monitoring of data compliance with privacy and quality standards										
	UR28	Implement data anonymization techniques to protect sensitive information										
	UR29	Develop a system for tracking data usage, sharing metrics, and maintaining version histories to ensure data traceability.										
		Use Case 2 and 6 TOTAL	3	4	2	3	1	1	0	0	0	0

Figure 16 Use case 2 and 6 functional requirements categorized by type

Use Case	UR Nº	Description	Technical	Performance	Reporting	User	Privacy	Security	Infrastructure	Access	Management	Collaboration
Use Case 3	UR30	Develop a centralized data repository for storing integrated health research data and implement data integration workflows to automate data ingestion from various sources										
	UR31	Implement APIs for seamless data sharing with external partners										
	UR32	Develop user-friendly data visualization tools										
	UR33	Develop a comprehensive access control system to manage user permissions										
		Use Case 3 TOTAL	1	0	0	1	0	0	1	1	0	0

Figure 17 Use case 3 functional requirements categorized by type

Use Case	UR Nº	Description	Technical	Performance	Reporting	User	Privacy	Security	Infrastructure	Access	Management	Collaboration
Use Case 4	UR34	Define and adopt standardized formats for annotation labels	■									
	UR35	Automate conversion of existing datasets format into standardized formats				■						
	UR36	Enable seamless data exchange between systems and devices through easy-to-use APIs							■			
	UR37	Validate incoming data against predefined schemas and standards	■									
	UR38	Create a metadata repository for standardized formats and protocols	■									
	UR39	Optimize edge computing for on-device data processing and filtering, eliminating the reliance on centralized systems	■									
	UR40	Automate deletion of outdated and expired data	■									
	UR41	Allow real-time processing of multimodal data	■									
	UR42	Integrate AI-driven insights and analytics into resource planning	■									
	UR43	Conduct real-time quality assessment of multimodal data	■									
	UR44	Provide intuitive interface for semi-automated annotation processes	■									
	UR45	Inject human knowledge during the annotation processes	■									
	UR46	Provide feedback to annotator for refining AI-generated annotations	■									
	UR47	Develop tools for real-time visualization of annotations	■									
	UR48	Generate reports on the annotation quality and consistency of data				■						
	UR49	Support metadata tagging in annotations to include context-specific information				■						
	UR50	Provide secure data-sharing infrastructure with encryption mechanisms	■									
	UR51	Enable datasets discoverability with AI-based brokering mechanisms and recommendation engines	■									
	UR52	Implement role-based access control for data access				■						
	UR53	Provide real-time anonymization of sensitive and personal data	■									
		Use Case 4 TOTAL	**15**	**0**	**0**	**4**	**0**	**0**	**1**	**0**	**0**	**0**

Figure 18 Use Case 4 Functional Requirements categorized by type

Use Case	UR Nº	Description	Technical	Performance	Reporting	User	Privacy	Security	Infrastructure	Access	Management	Collaboration
Use Case 5	UR54	Define a standard data set of parameters to be logged for lifetime monitoring of BESS	■									
	UR55	Develop and deploy necessary tools for collecting data based on defined standard dataset	■									
	UR56	Share necessary BESS data via the dedicated PLIADES Data Space (e.g. lifetime monitoring data, CAD, datasheets, Bill of Materials, etc.)				■						
	UR57	Manage data accessibility for different stakeholders within the platform				■						
	UR58	Develop dashboards for visualizing battery data life cycles	■									
	UR59	Develop analytics tools for life cycle data (e.g. battery state estimation and BESS predictive maintenance)	■									
	UR60	Create accurate predictions of battery status with the development and training of AI/ML models			■							
	UR61	Develop cross-domain AI models for federated applications										■
		Use Case 5 TOTAL	**4**	**0**	**1**	**2**	**0**	**0**	**0**	**0**	**0**	**1**

Figure 19 Use Case 5 Functional requirements categorized by type

The mapping result has been transferred to a graph showed below to present in a more visual way which are the main type of functional requirements identified by the different use cases during the internal discussions.

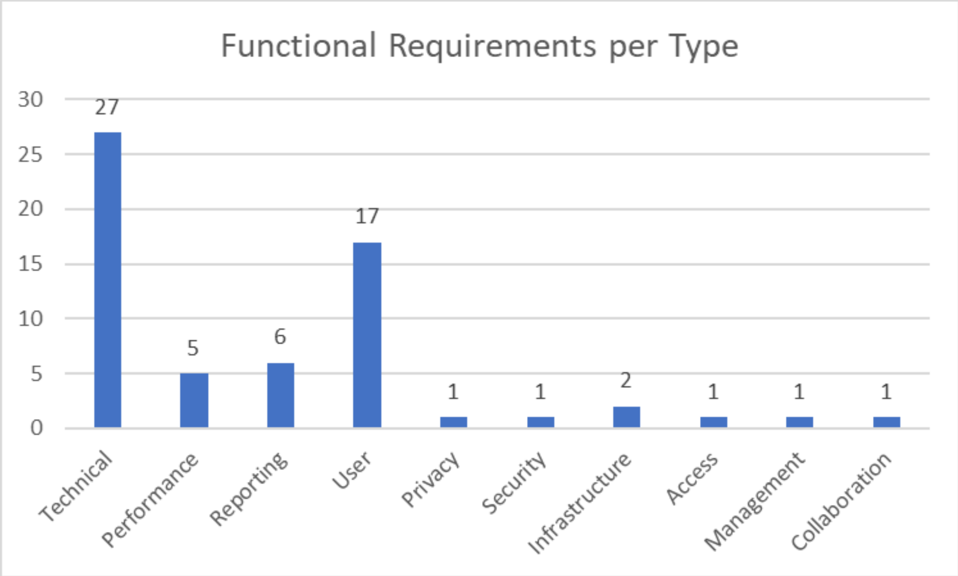

Figure 20 Bar plot for the complete set of functional requirements categorized by type

In the same way, similar categories were applied to the non-functional requirements:

- Performance
- Reporting
- User
- Privacy
- Security
- Infrastructure
- SupportStrategic
- Market Analysis
- Collaboration
- Legal

In the following figures, a categorization by type for each non-functional requirement is presented:

Use Case	UR Nº	Description	Performance	Reporting	User	Privacy	Security	Infrastructure	Support	Strategic	Market Analysis	Collaboration	Legal
Use Case 1	UR62	Utilize historical data to anticipate production bottlenecks and adjust processes proactively		X									
	UR63	Generate alerts for deviations in material quality based on predefined thresholds	X										
	UR64	Establish data governance policies to ensure data quality and consistency				X							
	UR65	Enable secure data sharing protocols across stakeholders, maintaining data sovereignty					X						
	UR66	Develop data anonymization techniques for sharing insights without compromising sensitive information				X							
	UR67	Develop tools for tracking data usage and auditing trails for accountability					X						
	UR68	Use advanced encryption techniques for data stored on cloud servers					X						
	UR69	Establish a framework for continuous data quality monitoring and improvement		X									
	UR70	Enable data-driven insights for regulatory reporting		X									
	UR71	Develop partnerships with technology companies to integrate DENNDATA into their solutions								X			
	UR72	Identify potential new sectors for DENNDATA expansion, such as healthcare or energy									X		
	UR73	Establish co-innovation programs with clients to develop new use cases for DENNDATA										X	
	UR74	Use predictive analytics to identify emerging market trends	X										
	UR75	Automate the generation of market reports based on data insights		X									
	UR76	Establish legal frameworks for co-development agreements with clients and partners											X
		Use Case 1 TOTAL	2	4	0	2	3	0	0	1	1	1	1

Figure 21 Use Case 1 non-functional requirements categorized by type

Use Case	UR Nº	Description	Performance	Reporting	User	Privacy	Security	Infrastructure	Support	Strategic	Market Analysis	Collaboration	Legal
Use Case 2 - 6	UR77	Define policies for sustainable data operations and archiving				X							
		Use Case 2 - 6 TOTAL	0	0	0	1	0	0	0	0	0	0	0

Figure 22 Use Case 2 and 6 non-functional requirements categorized by type

Use Case	UR Nº	Description	Performance	Reporting	User	Privacy	Security	Infrastructure	Support	Strategic	Market Analysis	Collaboration	Legal
Use case 3	UR78	Provide training programs on new data sharing protocols and tools							X				
	UR79	Develop training programs on data privacy and security protocols							X				
		Use Case 3 Total	0	0	0	0	0	0	2	0	0	0	0

Figure 23 Use Case 3 non-functional requirements categorized by type

Use Case	UR Nº	Description	Performance	Reporting	User	Privacy	Security	Infrastructure	Support	Strategic	Market Analysis	Collaboration	Legal
Use case 4	UR80	Define and adopt standardized data formats			■								
	UR81	Define and adopt standardized metadata formats			■								
	UR82	Monitor resource usage to optimize data operations			■								
	UR83	Ensure effective and scalable data operations			■								
	UR84	Implement cloud-based data operations for cost effectiveness						■					
	UR85	Train annotators on the use of the automated tools							■				
	UR86	Define and establish numerical metrics to evaluate annotation quality and consistency			■								
	UR87	Ensure compliance with privacy regulations in data-sharing activities			■								
	UR88	Define data access policies to control data sharing			■								
	UR89	Manage data sharing agreements and permissions through user-friendly interfaces			■								
	UR90	Continuously monitor and log data-sharing activity			■								
		Use Case 3 Total	0	0	9	0	0	1	1	0	0	0	0

Figure 24 Use case 4 non-functional requirements categorized by type

The mapping result has been transferred to a graph showed below to depict in a more visual way which are the main type of non-functional requirements identified by the different use cases during the internal discussions.

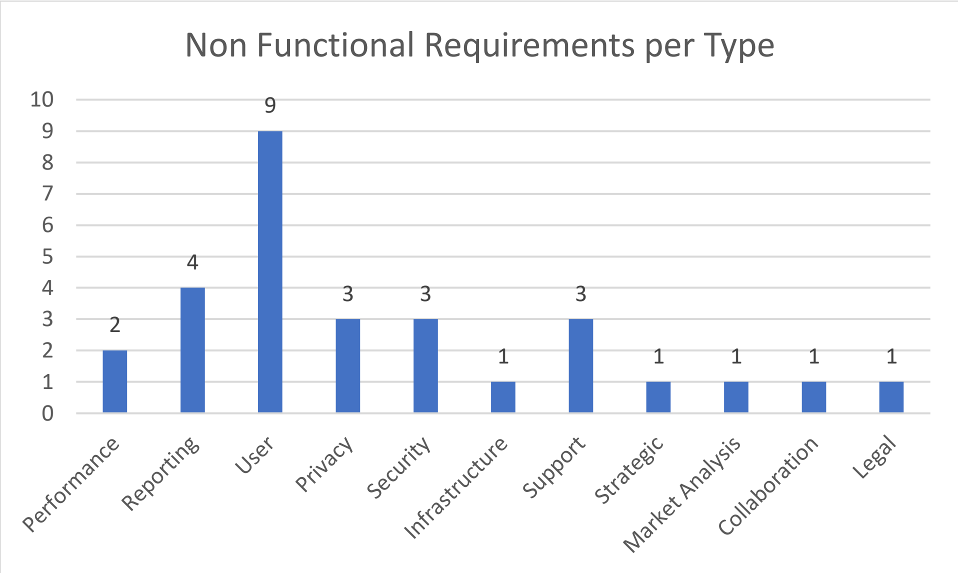

Figure 25 Bar plot for the global Non-functional requirements categorized per type

When analysing the functional requirements per use case and areas in which the implementation of the PLIADES framework will impact, analysis have been performed taking into account the following areas:

- Data Management
- Quality Control
- Management
- User Management
- Security
- Production
- Production Control
- Technical Support
- IT
- Design / Engineering

- Logistics
- Customer Relationships
- Data Analysis

Use Case	UR Nº	Description	Data management	Quality Control	Management	User Management	Security	Production	Production Control	Technical Support	IT	Design /Engineering	Logistics	Customer relationships	Data Analysis
Use Case 1	UR1	Implement automated quality checks based on real-time machine data													
	UR2	Enable flexible machine settings adjustments based on varying production requirements													
	UR3	Provide adaptive control for material specifications based on data-driven insights													
	UR4	Integrate quality data from previous production runs to improve current batch consistency													
	UR5	Facilitate collaborative adjustments between operators and quality managers for improved outcomes													
	UR6	Use predictive algorithms to adjust machine parameters before defects occur													
	UR7	Enhance data visualization tools for real-time monitoring of quality metrics													
	UR8	Dynamically adjust production schedules to account for material and product specification changes													
	UR9	Integrate human-in-the-loop methodologies to refine AI algorithms													
	UR10	Automate compliance checks for data access requests													
	UR11	Provide users with real-time compliance notifications for data operations													
	UR12	Implement pilot projects with early adopters in new markets													
	UR13	Develop training programs to help clients utilize DENNDATA for new business opportunities													
	UR14	Leverage AI to explore cross-industry data opportunities													
	UR15	Enable real-time collaboration tools for stakeholders working on new projects													
		Use Case 1 TOTAL	1	4	2	1	0	3	2	0	0	0	1	1	0

Figure 26 Use Case 1 Functional Requirements categorized per area

Use Case	UR Nº	Description	Data management	Quality Control	Management	User Management	Security	Production	Production Control	Technical Support	IT	Design /Engineering	Logistics	Customer relationships	Data Analysis
Use Case 2 -6	UR16	Implement compression techniques to optimize data transmission													
	UR17	Optimize edge computing for onboard filtering processes													
	UR18	Automate identification and discard of noisy and low-quality multimodal data													
	UR19	Develop semi-Automated methods for multimodal data quality estimation													
	UR20	Automate metadata tagging during data collection													
	UR21	Increase data discoverability through context-aware metadata													
	UR22	Automate analysis for cultural issues of human tracking data													
	UR23	Automate annotation processes for multimodal data													
	UR24	Implement compression and filtering techniques to reduce data storage needs													
	UR25	Automate identification and discard of redundant data													
	UR26	Implement advanced encryption protocols for all data													
	UR27	Automate monitoring of data compliance with privacy and quality standards													
	UR28	Implement data anonymization techniques to protect sensitive information													
	UR29	Develop a system for tracking data usage, sharing metrics, and maintaining version histories to ensure data traceability.													
		Use Case 2 and 6 TOTAL	7	5	0	0	2	0	0	0	0	0	0	0	0

Figure 27 Use case 2 and 6 Functional Requirements categorized per area

Use Case	UR Nº	Description	Data management	Quality Control	Management	User Management	Security	Production	Production Control	Technical Support	IT	Design /Engineering	Logistics	Customer relationships	Data Analysis
Use Case 3	UR30	Develop a centralized data repository for storing integrated health research data and implement data integration workflows to automate data ingestion from various sources													
	UR31	Implement APIs for seamless data sharing with external partners													
	UR32	Develop user-friendly data visualization tools													
	UR33	Develop a comprehensive access control system to manage user permissions													
		Use Case 3 TOTAL	0	0	0	0	1	1	1	1	0	0	0	0	0

Figure 28 Use Case 3 Functional Requirements categorized per area

Use Case	UR Nº	Description	Data management	Quality Control	Management	User Management	Security	Production	Production Control	Technical Support	IT	Design /Engineering	Logistics	Customer relationships	Data Analysis
Use Case 4	UR34	Define and adopt standardized formats for annotation labels													
	UR35	Automate conversion of existing datasets format into standardized formats													
	UR36	Enable seamless data exchange between systems and devices through easy-to-use APIs													
	UR37	Validate incoming data against predefined schemas and standards													
	UR38	Create a metadata repository for standardized formats and protocols													
	UR39	Optimize edge computing for on-device data processing and filtering, eliminating the reliance on centralized systems													
	UR40	Automate deletion of outdated and expired data													
	UR41	Allow real-time processing of multimodal data													
	UR42	Integrate AI-driven insights and analytics into resource planning													
	UR43	Conduct real-time quality assessment of multimodal data													
	UR44	Provide intuitive interface for semi-automated annotation processes													
	UR45	Inject human knowledge during the annotation processes													
	UR46	Provide feedback to annotator for refining AI-generated annotations													
	UR47	Develop tools for real-time visualization of annotations													
	UR48	Generate reports on the annotation quality and consistency of data													
	UR49	Support metadata tagging in annotations to include context-specific information													
	UR50	Provide secure data-sharing infrastructure with encryption mechanisms													
	UR51	Enable datasets discoverability with AI-based brokering mechanisms and recommendation engines													
	UR52	Implement role-based access control for data access													
	UR53	Provide real-time anonymization of sensitive and personal data													
		Use Case 4 TOTAL	7	2	0	0	0	0	0	0	11	0	0	0	0

Figure 29 Use Case 4 Functional Requirements categorized per area

Use Case	UR Nº	Description	Data management	Quality Control	Management	User Management	Security	Production	Production Control	Technical Support	IT	Design /Engineering	Logistics	Customer relationships	Data Analysis
Use Case 5	UR54	Define a standard data set of parameters to be logged for lifetime monitoring of BESS													
	UR55	Develop and deploy necessary tools for collecting data based on defined standard dataset													
	UR56	Share necessary BESS data via the dedicated PLIADES Data Space (e.g. lifetime monitoring data, CAD, datasheets, Bill of Materials, etc.)													
	UR57	Manage data accessibility for different stakeholders within the platform													
	UR58	Develop dashboards for visualizing battery data life cycles													
	UR59	Develop analytics tools for life cycle data (e.g. battery state estimation and BESS predictive maintenance)													
	UR60	Create accurate predictions of battery status with the development and training of AI/ML models													
	UR61	Develop cross-domain AI models for federated applications													
		Use Case 5 TOTAL	**2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**4**	**0**	**0**	**2**

Figure 30 Use Case 5 Functional Requirements categorized by area

When analysing the non-functional requirements per use case and areas in which the implementation of the PLIADES framework will impact, analysis have been performed taking into account the following areas:

- Data Management
- Management
- Production
- Production Control
- Human Resources
- IT
- Marketing
- Legal

Use Case	UR Nº	Description	Data management	Management	Production	Production Control	Human Resources	IT	Marketing	Legal
Use Case 1	UR62	Utilize historical data to anticipate production bottlenecks and adjust processes proactively								
	UR63	Generate alerts for deviations in material quality based on predefined thresholds								
	UR64	Establish data governance policies to ensure data quality and consistency								
	UR65	Enable secure data sharing protocols across stakeholders, maintaining data sovereignty								
	UR66	Develop data anonymization techniques for sharing insights without compromising sensitive information								
	UR67	Develop tools for tracking data usage and auditing trails for accountability								
	UR68	Use advanced encryption techniques for data stored on cloud servers								
	UR69	Establish a framework for continuous data quality monitoring and improvement								
	UR70	Enable data-driven insights for regulatory reporting								
	UR71	Develop partnerships with technology companies to integrate DENNDATA into their solutions								
	UR72	Identify potential new sectors for DENNDATA expansion, such as healthcare or energy								
	UR73	Establish co-innovation programs with clients to develop new use cases for DENNDATA								
	UR74	Use predictive analytics to identify emerging market trends								
	UR75	Automate the generation of market reports based on data insights								
	UR76	Establish legal frameworks for co-development agreements with clients and partners								
		Use Case 1 TOTAL	**2**	**4**	**2**	**4**	**0**	**0**	**2**	**1**

Figure 31 Use Case 1 non-functional requirements categorized per area

Use Case	UR Nº	Description	Data management	Management	Production	Production Control	Human Resources	IT	Marketing	Legal
Use Case 2 - 6	UR77	Define policies for sustainable data operations and archiving								
		Use Case 2 - 6 TOTAL	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**

Figure 32 Use Case 2 and 6 non-functional requirements categorized per area

Use Case	UR Nº	Description	Data management	Management	Production	Production Control	Human Resources	IT	Marketing	Legal
Use case 3	UR78	Provide training programs on new data sharing protocols and tools								
	UR79	Develop training programs on data privacy and security protocols								
		Use Case 3 Total	**0**	**0**	**0**	**0**	**2**	**0**	**0**	**0**

Figure 33 Use Case 3 non-functional requirements categorized per area

Use Case	UR Nº	Description	Data management	Management	Production	Production Control	Human Resources	IT	Marketing	Legal
Use case 4	UR80	Define and adopt standardized data formats	X							
	UR81	Define and adopt standardized metadata formats	X							
	UR82	Monitor resource usage to optimize data operations	X							
	UR83	Ensure effective and scalable data operations						X		
	UR84	Implement cloud-based data operations for cost effectiveness						X		
	UR85	Train annotators on the use of the automated tools					X			
	UR86	Define and establish numerical metrics to evaluate annotation quality and consistency	X							
	UR87	Ensure compliance with privacy regulations in data-sharing activities	X							
	UR88	Define data access policies to control data sharing	X							
	UR89	Manage data sharing agreements and permissions through user-friendly interfaces						X		
	UR90	Continuously monitor and log data-sharing activity	X							
		Use Case 3 Total	**7**	**0**	**0**	**0**	**1**	**3**	**0**	**0**

Figure 34 Use case 4 non-functional requirements categorized per area

Finally, the mapping result has been transferred to a graph showed below to depict in a more visual way where the needs identified by the 6 use cases are grouped. This mapping has also been performed for functional in Figure 35 and non-functional requirements in Figure 36.

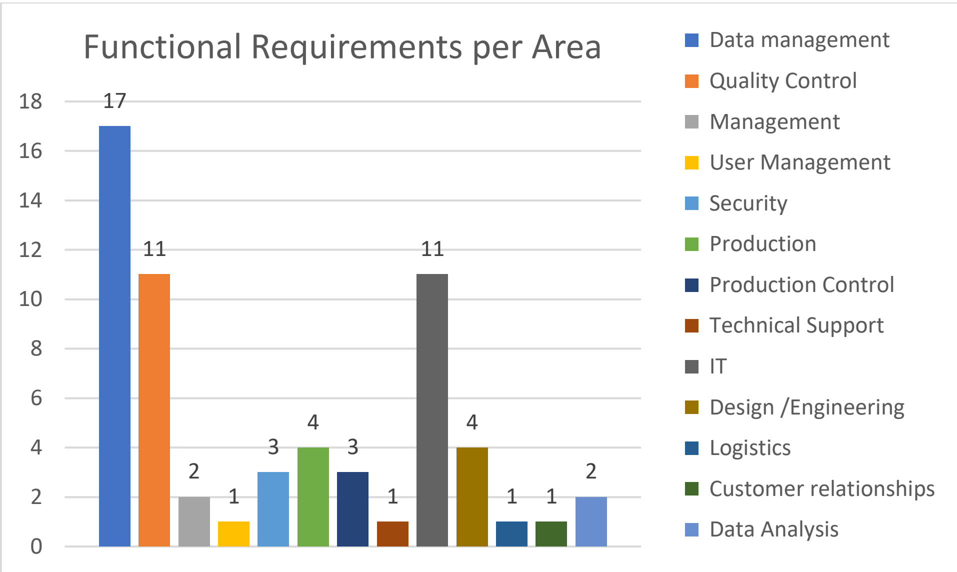

Figure 35 Functional Requirements categorized per area.

When analysing the use cases' functional requirements, the data management and the Quality Control and IT are the three most important areas where PLIADES could solve user needs. Despite the different sectors of each use case and the different scenarios they are facing, a proper data management aligned with the current European legislation is the most relevant. On the other hand, Management in terms of policies and "behaviours" while exchanging data is as relevant as providing a scalable IT infrastructure and resources to enhance a seamless operation between Data spaces.

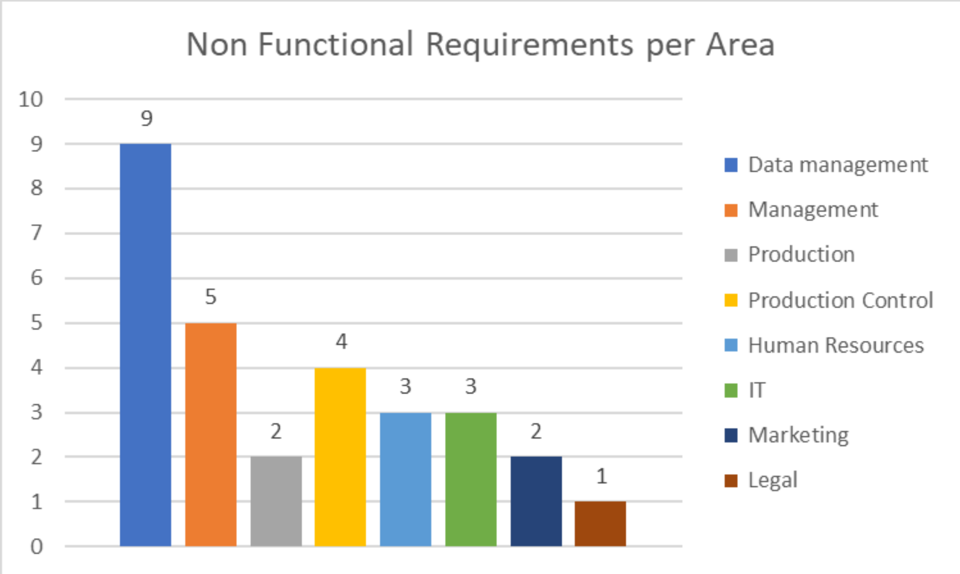

Figure 36 Non-Functional Requirements categorized per area

When analysing the use cases' non-functional requirements, as it happened with the functional requirements, the data management and management itself are the two most important areas where PLIADES could solve user needs. On the other hand, Production control is the third most important area as a result of the specifications identified in the industrial use cases that points out the relevance of applying Data Spaces technologies to boost the Industry 5.0 together with the Data Space Economy.

Finally, and by applying a high-level technical categorization on the Functional requirements, these are the main results:

- **Data Processing and Management**: includes requirements such as the implementation of compression techniques for data transmission (UR16), optimization of edge computing (UR17, UR39), automation of the identification and discarding of noisy data (UR18, UR25), methods for quality estimation of multimodal data (UR19), automation of metadata tagging (UR20), and the development of metadata repositories (UR38). It also includes the need to convert data to standardized formats (UR35). These requirements relate to storage technologies, processing, data transmission, and metadata management
- **Analytics and Artificial Intelligence:** encompasses the use of predictive algorithms (UR6, UR60), the integration of human knowledge into algorithms (UR9, UR45, and UR46), the generation of battery status predictions (UR60), and the development of AI models (UR61). In this area, machine learning, deep learning and predictive analytics technologies are used.
- **Quality and Control:** focuses on the implementation of automated quality controls (UR1), integration of quality data from previous productions (UR4), real-time monitoring of quality metrics (UR7) and quality assessment of multimodal data (UR43). Sensor technologies, real-time data analysis and quality control systems are applied here.
- **Visualization and Interaction:** includes the development of data visualization tools (UR7, UR32, UR47, and UR58), the implementation of intuitive interfaces (UR44), real-time collaboration tools (UR15) and the development of control panels (UR58). The technologies involved are user interface (UI) design, data visualization, collaboration tools and dashboards.
- **Interoperability and Connectivity:** Refers to the implementation of APIs for data sharing (UR31, UR36), the definition of standardized formats (UR34, UR80, UR81) and support for data exchange between systems (UR36). It relates to APIs technologies, data standards and communication protocols.
- **Access and Permissions Management:** Covers the development of access control systems (UR33, UR52, and UR57), the definition of data access policies (UR88) and the management

of sharing agreements (UR89). Identity and access management (IAM) technologies and security policies are used.
- **Process Automation:** includes automation of annotation processes (UR23), data format conversion (UR35), deletion of obsolete data (UR40), and analysis of cultural aspects in human-tracked data (UR22). Technologies include robotic process automation (RPA), process orchestration, and automation tools.

On the other hand, by applying the same high level technical categorization on the Non-Functional requirements, these are the main results:

- **Security and Privacy:** Includes requirements such as the definition of data governance policies (UR64), secure data sharing protocols (UR65), data anonymization techniques (UR66), data auditing (UR67), data encryption (UR68), compliance with privacy regulations (UR87), and data access policies (UR88). Relates to cybersecurity technologies, data privacy, and regulatory compliance management.
- **Data Quality:** Covers continuous monitoring of data quality (UR69), use of historical data to anticipate problems (UR62), generation of alerts for quality deviations (UR63), definition of metrics for annotation quality (UR86), and data validation with schemas and standards (UR37). Focuses on data quality management technologies and monitoring systems.
- **Performance and Scalability**: Comprises optimizing the use of resources (UR82) and ensuring scalability of operations (UR83). Relates to cloud infrastructure technologies and resource optimization.
- **Sustainability:** Covers the definition of policies for sustainable data operations (UR77). Focuses on technologies for energy efficiency and resource management.
- **Collaboration and Interoperability:** Includes the definition of standardized data formats (UR80, UR81), management of sharing agreements (UR89), and monitoring of data sharing activity (UR90). It relates to collaboration technologies and interoperability standards.
- **Training and Adoption:** Covers the development of training programs for the use of new tools (UR78, UR79, and UR85) and the training of annotators (UR85). Learning and training technologies are used.
- **Legal and Business Aspects:** Includes the establishment of legal frameworks for co-development agreements (UR76), the identification of new sectors (UR72), the establishment of co-innovation programs (UR73), and the generation of automated market reports (UR75).

5 Conclusions and next steps

The PLIADES project, involving 27 participants, encompasses diverse cross-sectoral use cases, reflecting a wide range of perspectives and expertise in defining user requirements. The goal of **D2.2 “User Requirements and Use Cases”** is to provide a comprehensive analysis of the identified user needs and establish detailed specifications for the PLIADES use cases. This deliverable captures the progress achieved through **T2.2 “Consolidation of user and system requirements”** and **T2.4 “Design of the Use Cases for the deployment of the multiple project data spaces”**.

These activities culminated in the definition of precise requirements, creating a foundation for the successful deployment and real-world demonstration of PLIADES tools. The document explored current business scenarios and workflows, followed by the development of future scenarios that aligned with the framework's objectives and addressed challenges faced by participating organizations in ensuring successful implementation.

The design of the PLIADES use cases and framework reflects the diverse needs of the demonstrators and highlights several critical aspects. The analysis of functional requirements indicates a strong emphasis on robust data management, quality control, and IT infrastructure. The framework is expected to prioritize efficient, secure, and compliant mechanisms for collecting, storing, processing, and sharing data in line with European regulations. Key functionalities include the implementation of automated quality controls using real-time machine data, the ability to adapt machine configurations flexibly based on production demands, and the integration of quality data from previous runs to improve batch consistency. Real-time data visualization, collaboration tools for operators and managers, and advanced analytical tools to support decision-making are considered essential. Additionally, the framework should facilitate automated data conversion into standardized formats, ensure seamless data exchange through APIs, and validate incoming data against predefined schemas and standards. Features such as a metadata repository for standardized formats and protocols, optimization of edge computing for on-device processing, and real-time multimodal data processing are critical. The integration of AI-driven insights for resource planning, secure data exchange mechanisms with encryption, and the development of interdomain AI models for federated applications are also essential components.

From a non-functional perspective, the focus on data management and policy underscores the need for a robust data governance strategy to ensure information quality and consistency. Leveraging historical data to anticipate production bottlenecks, generating alerts for material quality deviations, and complying with privacy regulations are pivotal for operational excellence. Establishing legal frameworks for co-development agreements, defining policies for sustainable data operations and archiving, and standardizing metadata formats contribute to a cohesive and reliable framework. The monitoring and optimization of resource use, scalability of operations, and implementation of cloud-based solutions to improve cost efficiency are essential considerations. Furthermore, training annotators on automated tools and defining numerical metrics to evaluate the quality and consistency of annotations are critical for maintaining high data standards.

The requirements outlined in this deliverable, along with the future scenarios for use case design, align with other outcomes in WP2, including the D2.3 “System Technical Specifications and PLIADES Framework Architecture.” The next steps for the project emphasize developing the data spaces infrastructure to enable secure and efficient data exchange across the six use cases, establishing standardized data formats, metadata, and annotation tags to ensure interoperability, and creating advanced tools for data quality management, annotation, and metadata handling. Security measures, including encryption, anonymization, and role-based access control, remain central to ensuring data privacy and compliance with GDPR and other regulations.

In summary, the implementation of the PLIADES framework requires a comprehensive and holistic vision that considers technical as well as **legal, ethical and social** aspects to address the evolving needs of stakeholders and foster innovation across sectors.

6 References

1. Bonilla, L., López Osa, M. J., Diaz-de-Arcaya, J., Torre-Bastida, A. I., & Almeida, A. (2024, August). Purity: a New Dimension for Measuring Data Centralization Quality. In Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing (pp. 8-14).
2. Zarate, G., Lopez Osa, M. J., Torre-Bastida, A. I., Iturraspe, U., Arjona, J., Navarro, B., & Gimeno, A. (2024, October). Evolution of Extract-Transform-Load (ETL) processes towards data product pipelines. In Proceedings of the 4th Eclipse Security, AI, Architecture and Modelling Conference on Data Space (pp. 25-32).
3. Nikolakopoulos, A., Chondrogiannis, E., Karanastasis, E., Osa, M. J. L., Aroca, J. A., Kefalogiannis, M., ... & Varvarigou, T. (2024, June). Scalable Data Profiling for Quality Analytics Extraction. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 177-189). Cham: Springer Nature Switzerland.
4. Aroca, J. A., Osa, M. J. L., Barturen, U. I., Siopidis, V., Votis, K., Nikolakopoulos, A., ... & Nestorakis, K. (2024, June). Boosting Data Monetisation with DATAMITE. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 33-46). Cham: Springer Nature Switzerland.
5. Towards an Advanced Artificial Intelligence Architecture Through Asset Administration Shell and Industrial Data Spaces https://link.springer.com/chapter/10.1007/978-3-031-57496-2_4 ESAIM 2023
6. Towards Standardized Manufacturing as a Service through Asset Administration Shell and International Data Spaces Connectors https://ieeexplore.ieee.org/abstract/document/9968592/ - IECON 2022
7. Towards an Asset Administration Shell scenario: a use case for interoperability and standardization in Industry 4.0 https://ieeexplore.ieee.org/abstract/document/9110410 NOMS 2020
8. Data Sovereignty for AI Pipelines: Lessons Learned from an Industrial Project at Mondragon Corporation https://ieeexplore.ieee.org/document/9796393
9. Urbieta, I., Nieto, M., García, M., & Otaegui, O. (2021). Design and implementation of an ontology for semantic labeling and testing: automotive global ontology (AGO). Applied Sciences, 11(17), 7782.

7 ANNEX I: Trial Handbook Chapter 2 / Business Requirements Questionnaire

The following section contains the "Business Requirements Questionnaire" that was filled-in by the use case participants, in order to collect the requirements of each scenario.

Preface

Each Use Case leader (in collaboration with the technology providers representatives), should complete the information requested in the different sections.

In this document, each Use Case representative should identify one (or two) business scenarios that represent the real situations in the context of the demonstrator

Each BUSINESS SCENARIO involves a number of specific processes. It should include the following inputs:

- Description of each business scenario within each use case
- Explanation of the current situation/scenario in the company
- Directives and regulations
- Description of the future scenario
- Business objectives
- Expected benefits
- Business indicators
- Drivers

1 Use Case Framework and Barriers

1.1 Legal Framework

Give a description of the legal framework related to the technology situation.

Please provide a relationship and description of the framework and the impact (personal data protection, labor security, quality, work conditions, etc.).

- Regulations and legislation (At International level, at EU level, at national level, at regional/legal level) applicable to the either the technology to be implemented / the company or sector or industry / the specific products and services to be tackled / the trial itself.
- Other related policies in place and in process of implementation.
- Standards and certifications. Are there any certifications applicable to the trial (technology, products or services)?
- Legal implications

It is important to summarise all the information in the tables below

(between 1 and 2 pages, not including the tables)

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101135988

Table 29- Sector/Industry Framework & Impact summary

SECTOR / INDUSTRY (Mention the sector/industry of the end-user)	**LEGISLATION / REGULATIONS / STANDARDS / CERTIFICATIONS** (related to the sector/industry and applicable to the trial)	**LEVEL** (International, EU, national, regional, local)	IMPACT (Positive/Negative) Indicate how the legislation/Standards affect the sector/industry	DESCRIPTION Describe and justify the impact that the listed legislation/standards will have when implementing the trial.	BARRIERS List the barriers that can appear as a result of the legislation related.

Table 30- Technologies Framework & Impact summary

TECHNOLOGIES (List the technologies involved in the Trial)	**LEGISLATION / REGULATIONS / STANDARDS / CERTIFICATIONS** (related to any of the technologies to be implemented)	**LEVEL** (International, EU, national, regional, local)	IMPACT (Positive/Negative) Indicate how the legislation/Standards will affect the technology's implementation	DESCRIPTION Describe and justify the impact that the listed legislation/standards will have when implementing the technology of the trial.	BARRIERS List the barriers that can appear as a result of the legislation related.

Table 31- Products/Services Framework & Impact summary

PRODUCTS / SERVICES (List the products and services involved in the Trial)	**LEGISLATION / REGULATIONS / STANDARDS / CERTIFICATIONS** (related to any of the previous products and services)	**LEVEL** (International, EU, national, regional, local)	IMPACT (Positive/Negative) Indicate how the legislation/Standards will affect the product or service	DESCRIPTION Describe and justify the impact that the listed legislation/standards will have when implementing the technology of the trial.	BARRIERS List the barriers that can appear as a result of the legislation related.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101135988

1.2 Social Framework

Give a description of the social framework that characterizes the current situation of the company's sector and the technology related to the trial. List and describe all social aspects that affect the activity performed at your factory, describing the impact and identifying their causes.

- Social relations
- Gender issues
- Equality and diversity
- Globalisation
- Privacy
- Security
- Health
- Sustainability
- Culture
- Tradition and customs
- Etc…

(Between 1 and 2 pages)

1.3 Economic Framework

Give a description of the economic framework that characterizes the current situation of the company's sector and the technology related to the trial. List and describe all economic aspects that affect the activity performed at your factory, describing the impact and identifying their causes.

(between 1 and 2 pages)

2 Business Scenario

These are working examples of your business processes. They provide an opportunity to walk through the process as a user to better understand how the system works, or should work, and the function of the new solution in that process.

Pilot representative should identify X Business Scenarios (as many as you need), which will be characterized by their Business Objectives and Y Business Processes (as many as you need).

Fill in the following tables.

Table 32- Business Scenario

<table>
<tr><td>Business Scenario / Use case
Acronym / Name</td><td colspan="4"></td></tr>
<tr><td>Location</td><td colspan="4"></td></tr>
<tr><td>Number of persons involved</td><td>End user</td><td></td><td>Technology provider</td><td></td></tr>
<tr><td>Responsible/coordinator</td><td colspan="4"></td></tr>
<tr><td>Activity area</td><td colspan="4">Health, manufacturing line, mobility, quality, etc.</td></tr>
</table>

2.1 Current Business Scenario

Describe in detail the company's current way to perform the activities the technology is focused into (business scenario/use case). List the objectives that your company is attempting to achieve due to the use case.

Enrich the understanding with a storyline and graphic support if possible.

(between 2 and 3 pages)

2.2 Future Business Scenario

Describe in detail the company's current way to perform the activities the technology is focused into (business scenario/use case). List the objectives that your company is attempting to achieve due to the use case.

Enrich the understanding with a storyline and graphic support.

(between 2 and 3 pages)

2.3 Business Process

A business process is an activity or set of activities that will accomplish a specific organizational goal. It will be needed to define and describe at least one Business Processes per each Business Scenario.

2.3.1 Business Process In Detail

Describe at least one business process that will be affected by the Business Scenario/ Use case. Detail the current business process and the expected evolution due to Business Scenario/Use Case.

(Between 1 and 2 pages)

2.3.2 Business Objectives

Regarding the previous defined business process define a set of business objectives that the company expects to achieve. Justify the alignment of the new scenario with the expected benefits set as business objectives.

Once detailed, complete the summary table provided below.

(Between 1 and 2 pages) This limit doesn´t include the table.

This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101135988

Table 33- Business Process: Business Objective Summary

BUSINESS OBJECTIVE *List the Business objectives expected for the Business Scenario/Use Case*	DESCRIPTION *Give a short description about the Business Objective*	**IMPACT** *Indicate the expected impact of the business objective for the company*	EFFECT IN VALUE[1] *Rate from 1 to 5 (being 1 no significant impact and 5 very high impact)*	
			Cost	
			Efficiency	
			Quality	
			Flexibility	
			Innovation	
			Sustainability	
			Cost	
			Efficiency	
			Quality	
			Flexibility	
			Innovation	
			Sustainability	

These six categories of effect proposed, have also been proposed in European Projects like EUR3KA, XMANAI or AUTOTWIN, where the aspects at the business process level in which an improvement is expected to occur are:

Cost	*ICT capabilities can contribute to cost reduction in various ways: Automation and standardization of tasks, qualify workers faster and better, inventory and stock management improve and integrating the suppliers into its customer´s ICT.*
Efficiency	*ICT capabilities can contribute to improve efficiency in various ways: Improve availability and efficiency of machines, enabling easy access of maintenance employees to relevant information, facilitating identification and analysis of problems, shortening the response time to malfunctions, optimization of the machine´s activity scheduling and automating decisions regarding maintenance and production activities.*
Quality	*ICT capabilities can contribute to improve quality in various ways: Supporting the creation of the characteristics of the product or service, understanding the requirements that the product or service should fulfil, increase the feet to the specification of the product or service (decrease errors), improve the communication of the involved parties in the creation and delivery of the product or service, making information about the customer´s order available from initiation to completion, facilitating transparency and early identification of deviation from desire outcome, adapting to the changing need of customer's and the constrains of suppliers, facilitating better understanding of customer´s needs and generating inside into de customer´s tacit as well as explicit needs and requirements.*
Flexibility	*ICT capabilities can contribute to improve flexibility by product and service customization to adapt to market changes and customers preferred options at designed stage.*
Innovation	*ICT capabilities can contribute to improve innovation in various ways: Common learning and knowledge sharing and facilitating information, integration and accessibility among supply chain partners.*
Sustainability	*ICT capabilities can contribute to improve sustainability by reducing carbon footprint in three ways: Reduced paper usage, reduction of energy consumption and reduce fuel consumption in distribution.*

3 BUSINESS INDICATORS

Define tangible and intangible business indicators that can help to measure the achievement of the business objectives. (Reduction of time in manufacturing, quality enhancement, reduction of accidents/business interruptions, increase in benefits, reduction of costs, etc.).

Fill in the table provided below

Table 34- Business Process: Business Indicators Summary

ID	**BUSINESS Indicators** *List the Business objectives expected for the Business Scenario/Use Case*	**DESCRIPTION** *Give a detailed description of the indicators*	**Unit***	**Current value**	**Future expected value**	**Expected date of achievement****
1						
2						
3						
4						
5						

** Provide the units, in which you measure the value of the defined indicator (%, people, €, Si, etc.).*

*** During the implementation / end of the implementation / before 6 months after implementation / before 12 months after implementation / before 24 months after implementation / more than 24 months after implementation*

4 ACTORS INVOLVED

List and describe all persons whose activities will be affected, directly or indirectly, by the implementation of the use case.

Fill in the table provided below

Table 35- Business Process: Actors involved summary

ACTOR *(Blue Collar Workers, Manager, Coordinator, clients, provider, etc.)*	**BUSINESS AREA** *(marketing, administration, manufacturing, etc.)*[10]	**TYPE OF IMPACT** **(Direct or Indirect)**	DESCRIPTION OF THE IMPACT *Detail the expected impact of the Business Scenario/Use Case for each actor in their Business Area*

[10] Choose the adequate from the following list:

- Management
- Manufacturing
- Purchasing
- Marketing
- Sales
- Technical Support
- Accounting
- Customer Relationships
- Warehousing
- Others: (Indicate)

5 EXPECTED BENEFITS

Describe the benefits expected at a general level for the industry with the implementation of the user case technology.

Fill in the table below

Table 36- Business Process: Expected benefits summary

BENEFIT *List the benefits identified in the Business Scenario/Use Case*	DESCRIPTION *Describe the related benefits for the company*	**COMPANY IMPACT** *Justify the impact for the company*	**POTENTIAL INDUSTRY IMPACT**	
			Impact (High/Medium/Low)	*Justify the impact for the industry of the listed benefits*

6 BUSINESS REQUIREMENTS

Business requirements are statements of what a system should do rather than how it should do it. They are instructions describing the functions the system should provide and the characteristics the solution should have. They answer the question: ***"what does the business want to do?"***

When developing or implementing a new Technology solution, business requirements serve a number of important purposes. For example:

- *Help a business understand and articulate what they are looking for and provide a framework for them to make an informed decision*
- *Help manage the scope of a Technology development or purchase*
- *Provide a mechanism to communicate to a technology service provider what the solution needs to do to satisfy the business' needs*
- *Inform cost and product pricing decisions*

Aligned with the business objectives described above, section **2.3.2.**, list and define the specific business processes modifications that will permit the achievement of the objectives.

Fill in the table provided bellow, taking into account the glossary (add as many rows as needed)

(Look at example table provided for understanding)

NOTE: Each business objective should result in at least 10 business requirements.

Table 37- Business Process: Business requirements summary

No.	BUSINESS OBJECTIVE	REQUIREMENT *Provide a short description of the requirement*	AREA[1]	SUB HEADING[2]	FUNCTIONALITY[3]	PRIORITY[4]
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

[1] **Indicate to which area the requirement is addressed to:** • **Management** • **Marketing** • **Production** • **Production Control** • **Financial** • **Quality Control** • **Logistics** • **Human Resources** • **Sales & Purchases** • **Maintenance** • Others: (Indicate)	[2] **Choose the adequate from the following list:** • **User Requirements** • **Technical Requirements** • **Infrastructure Requirements** • **Reporting Requirements** • **Access Requirements** • **Security Requirements** • **Privacy Requirements** • **Performance Requirements** • **Support Requirements** • Others: (Indicate)	[3] **Choose one of the following:** • **Functional Requirements: These describe how a solution should function from an end user´s perspective. They describe the features and functions of the system required by the user.** • **Non-Functional Requirements**: These describe the operational characteristic of the system. These could relate to availability, accessibility, performance, scalability, auditability, etc. [4] Choose the adequate from the following list: **Critical**: These are those types of requirements without which the business objective is achievable **Preferred**: These requirements are those without which, the business objectives are achievable but not in the most efficient and effective way. **Optional**: These requirements although desired do not affect the business objective defined.

***Example**: Textile+ is a clothing manufacturing company. In order to shield their position in the market in front of the emerging market competitors, they set the following business objectives:*

- *Reduce time to market.*

NOTE: For the sake of understanding, requirements presented are at a general level. TRIAL Business Requirements should be focus at a more specific level.

No.	BUSINESS OBJECTIVE	REQUIREMENT	AREA[1]	SUB HEADING[2]	FUNCTIONALITY[3]	PRIORITY[4]

		Provide a short description of the requirement				
1	**Reduce time to market**	Operators should have an easier access to the tools they need	Production	User	Functional	Preferred
2		Users should be better trained for their jobs	Production	Support	Non-Functional	Preferred
3		Production line machinery will be desired to be renovated	Production	Infrastructure	Non –Functional	Optional
4		Goods needed for production must arrive more frequently	Sales & Purchase	Other: Provider	Non-Functional	Critical
5		Quality control reporting timing must be reduced	Quality control	Reporting	Non-Functional	Optional
6		Distribution routes should be adapted/optimized	Logistics	Performance	Non-functional	Preferred
7		Must track store by BIN and warehouse location	Logistics	Infrastructure	Functional	Critical
8		Protocols for production control should not constrain the production rate	Production Control	Support	Non-Functional	Preferred
9		Provide graphic display/reports of expected work load at work station compared to capacity at work station	Production Control	User	Functional	Preferred
10		Maintenance activities must be adapted to the new production rate	Maintenance	Support	Non-Functional	Critical